

ADA036538

FINAL REPORT  
FOR LINE ITEM A002:

DEMONSTRATION OF WESTINGHOUSE AUTOMATIC  
CUEING TECHNIQUES USING NVL IMAGERY

May 23, 1976

Contract DAAG53-75-C-0225

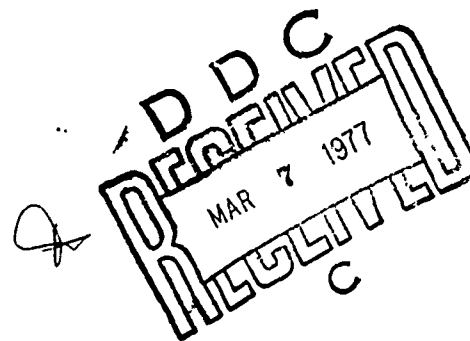
Prepared for the

NIGHT VISION LABORATORY  
U. S. Army Electronics Command  
Ft. Belvoir Virginia 22060

By

WESTINGHOUSE DEFENSE AND ELECTRONICS SYSTEMS CENTER  
Systems Development Division  
Baltimore, Maryland 21203

Distribution Unlimited



UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)


REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
	6	9
4. TITLE (and Subtitle) Final Report For Line Item A002: Demonstration of Westinghouse Automatic Cueing Techniques Using NVL Imagery,		5. TYPE OF REPORT & PERIOD COVERED Final Rept.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) 10 G. E. Tisdale T. P. Truitt		8. CONTRACT OR GRANT NUMBER(s) 15 DAAG53-75-C-0225 new
9. PERFORMING ORGANIZATION NAME AND ADDRESS Westinghouse Defense & Electronic Systems Center Systems Development Division Baltimore, Md. 21203		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 11 23 May 76
11. CONTROLLING OFFICE NAME AND ADDRESS Night Vision Laboratory U.S. Army Electronics Command Ft. Belvoir, Va. 22060		12. REPORT DATE May 23, 1976
		13. NUMBER OF PAGES 89
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 12 91p.		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Distribution Unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) D D C RECEIVED MAR 7 1977 REGULATED C.		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Automatic Target Cueing      Statistical Test      Detection Target Recognition      Recognition Features FLIR Sensor      Two-Layer Classifier Digital Image Processing      Decision Boundaries		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) See Page ii.		

## FOREWARD

This final report is submitted in compliance with Item A002, Contract DAAG53-75-C-0225, to the U.S. Army Electronics Command, Night Vision Laboratory, Ft. Belvoir, Virginia, by the Systems Development Division, Westinghouse Electric Corporation, Baltimore, Maryland, 21203.

The project engineer at the Night Vision Laboratory is Mr. G. David Singer.

The Westinghouse engineering team consists of G. E. Tisdale, program manager, T. P. Truitt, and D. T. Bissell.

Approved:   
G. David Singer  
Project Engineer

REVISION	<input checked="" type="checkbox"/>
DEFINITION	<input type="checkbox"/>
DESIGN	<input type="checkbox"/>
TESTING	<input type="checkbox"/>
ON LINE	<input type="checkbox"/>
ANALYSIS	<input type="checkbox"/>
APPROVAL	<input type="checkbox"/>
REVIEW	<input type="checkbox"/>
FINAL	<input type="checkbox"/>

## ABSTRACT

✓ This report describes the work and results of a study to establish the performance of existing digital image processing techniques on FLIR imagery supplied by NVL. The image processor would form the basis for an automatic target cueing system to assist the human operator of a sensor system.

The study consisted of a statistical test, performed by computer simulation, including training and test phases. The target classes included truck, tank, and APC. Initial detection of targets scored in the 90% range. Depending upon image quality, the classification performance was in the 60% to 80% range. Using the same sensitivity setting, the false alarm rate was 20%. The exact setting, trading false alarm rate for detection rate, would depend upon the mission requirements.

It was noted that the number of samples was very limited, in view of the number of features used. Future efforts might include a larger data base. It was also suggested that the design of a compact automatic cueing system breadboard be started to keep pace with sensor hardware development.

↑

## TABLE OF CONTENTS

	Page
FORWARD	
ABSTRACT	
1.0 INTRODUCTION	1-1
2.0 DESCRIPTION OF WESTINGHOUSE AUTOMATIC CUEING TECHNIQUES	2-1
2.1 Preprocessor	2-3
2.1.1 Gradient Extraction	2-3
2.1.2 Gradient Maximizing	2-5
2.1.3 Subset Generation	2-5
2.1.4 Blob Detector	2-8
2.1.5 Texture Data	2-10
2.2 Final Processor	2-10
2.2.1 Blobs and Groups	2-12
2.2.2 Feature Generation	2-14
2.2.3 Recognition Algorithm	2-18
3.0 TEST PROGRAM USING NVL IMAGERY	3-1
3.1 The Data Base	3-1
3.2 Preparation of Imagery	3-4
3.3 The Image Processing Sequence	3-24
3.4 Training Program	3-27
3.5 Scoring of Training Set	3-31
3.6 Test Set Performance	3-41
3.7 Discussion	3-52
4.0 CONCLUSIONS AND RECOMMENDATIONS	4-1
5.0 REFERENCES	5-1

# LIST OF ILLUSTRATIONS

Figure No.	Title	Page No.
2-1	Automatic Cueing System	2-2
2-2	Steps in Automatic Cueing	2-2
2-3	Digital Image Processor	2-4
2-4	Gradient Extraction Process	2-6
2-5	Steps in the Preprocessing of an Image	2-7
2-6	Block Diagram - Blob Detector and Subset Generator - Operational Cycle	2-9
2-7	Final Processor	2-11
2-8	Blob Merging	2-13
2-9	Significance of Polarities Between Subsets	2-14
2-10	Recognition Algorithm - Block Diagram	2-19
2-11	Regional Boundary Sets B G, K, L	2-22
2-12	Classification Logic Flow	2-23
3-1	Sketch of Targets	3-3
3-2	A 50 x 50 Window Extracted from an Image	3-6
3-3	Photo Playbacks of 50 x 50 Images	3-12 thru 3-23
3-4	Two Examples of Degraded Samples	3-25
3-5	Data Flow Through Simulated Processor	3-28
3-6(a)	Training Set Scatterplots	3-32
3-6(b)	Training Set Scatterplots	3-33
3-6(c)	Training Set Scatterplots	3-34
3-7	Extraneous White Line in Image	3-40

# LIST OF TABLES

Table No.	Title	Page
2-I	False Alarm Rejection Criteria	2-20
2-II	Tie-Breaking Rules	2-24
3-I	List of Digital Tapes Available	3-2
3-II	List of Data Windows	3-8 thru 3-11
3-III	Number of Samples	3-26
3-IV	Training Set Results - Raw Scores	3-35
3-V	Training Set Results - Percentages	3-37
3-VI	Test Set Results - Raw Scores	3-42
3-VII	Test Set Results - Percentages	3-44
3-VIII	Training and Test Results - By Window	3-45 thru 3-48
3-IX	Sum of Test and Training Results	3-53
3-X	Interpreter Performance in Recognition	3-55

## 1.0 INTRODUCTION

The major business activity of the Westinghouse Systems Development Division consists of the development of sophisticated sensor systems for military requirements. The programs cover radar, IR and visual frequencies. In 1965, pattern recognition research was initiated within the Division to support these sensor programs. The objective of this research was the development of digital image processing and automatic recognition techniques and systems.

By 1970, a specific approach had been established for the extraction of useful information, such as target location and identity, from remote sensor images. The approach consists of the serial preprocessing of the digitized image samples, on a line-by-line basis, so as to extract certain key image features, and to reduce the data bandwidth by orders of magnitude. The results of the preprocessing operation are then operated on by a general-purpose processor, to locate and classify targets, or to perform map-matching between similar terrain images. The Westinghouse techniques for digital image processing are described in Section 2.0.

At about the same time, the military laboratories began to support this research for the specific application to the problem of "automatic target cueing". We might define automatic cueing as the use of automatic recognition devices to initiate appropriate audible or visual signals (cues) to assist the human interpreter in his evaluation of sensor images. The cueing system acts as an information filter on the sensor data, by selecting important events, by providing an audible alarm to attract the attention of the operator, and

then by providing visual indications of the target location and identity on his display. In 1971, Westinghouse began automatic cueing studies with the Naval Air Systems Command (Reference 1), with the Air Force Rome Air Development Center (Reference 2), and with the Army's Frankford Arsenal (Reference 3). In the latter program, a real-time demonstration breadboard cueing system was constructed, which is presently being tested with video-taped flight data.

In general, the results of these programs are very promising when compared with available performance data for human operators under realistic circumstances. It appears quite possible that the target acquisition performance of a helicopter pilot, for example, might be doubled with the use of automatic cueing devices.

In February, 1975, a presentation on Westinghouse cueing techniques was made to Mr. John Dehne and Dr. James Tegnalia of NVL. Following that meeting, Mr. Dehne indicated that NVL was preparing a data base of digitized images for an 875-line TV compatible FLIR sensor. He expressed an interest in the performance of the Westinghouse technique on this data base. The program described in this report provides an answer to that question.

The description of the techniques in Section 2.0 is followed by a detailed discussion of the test program, using the imagery supplied by NVL, in Section 3.0. Conclusions and recommendations are contained in Section 4.0, and References in Section 5.0.

## 2.0 DESCRIPTION OF WESTINGHOUSE AUTOMATIC CUEING TECHNIQUES

We define "automatic cueing" as the use of automatic recognition devices to initiate appropriate audible or visual signals or cues to assist the human interpreter in his evaluation of sensor images. As shown by Figure 2-1, the cueing system acts as an information filter on the sensor data, selecting images of importance, marking them with visual indications of target location, and providing audible alarms to attract the attention of the interpreter.

The sequences of operations carried out by an automatic cueing system is shown by Figure 2-2. The operations are performed over the entire image, although the figure examines only a small window of the FLIR display shown at the top. First; the image is digitized for use by the image processor. Preprocessing of the digitized data serves to reduce its bandwidth by retaining only the information necessary for automatic recognition. When recognition of desired targets has been accomplished, appropriate audible and visual cues are initiated. These cues will not only identify the target types, within the limitations of sensor resolution, but can also provide precise coordinates of their location in the image. A variety of target types can be accommodated simultaneously by the cueing system.

The core of the cueing system is the digital image processor. It is a hybrid system utilizing a high speed hardwired preprocessor, followed by a programmable processor (general-purpose computer) to generate features and employ the recognition logic. The preprocessor is provided as special-purpose hardware in order to achieve a high data input rate. The output data rate is greatly reduced (by at least a factor of five), permitting the flexibility available in a slower speed programmable processor for final

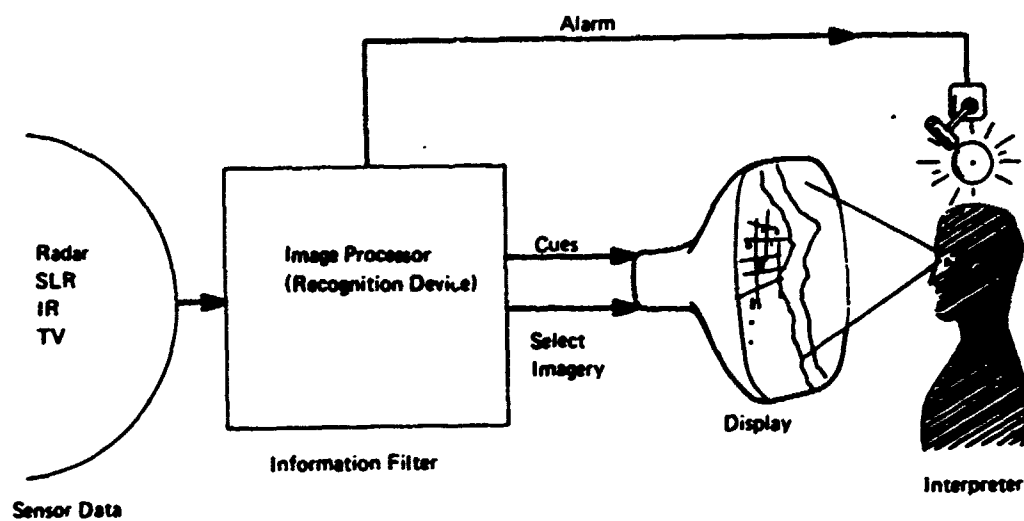


Figure 2-1. Automatic Cueing System

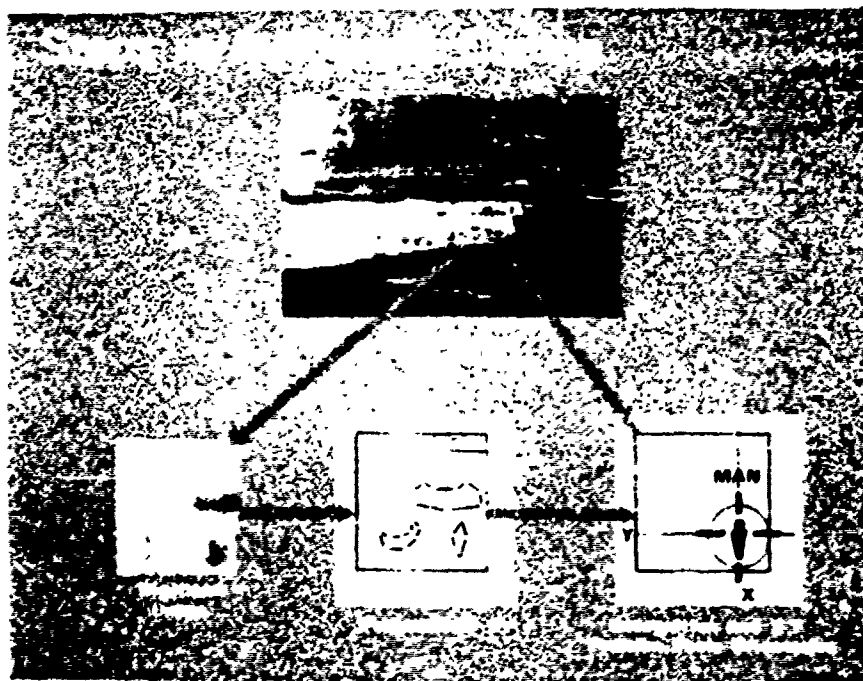


Figure 2-2. Steps in Automatic Cueing

target decisions. A block diagram of the image processor is shown by Figure 2-3. It should be noted that the image processor is a two-dimensional processor. The preprocessor contains 4 sets of single-line storage that "wrap around" to permit two-dimensional operations. Operation in both dimensions simultaneously provides greater noise rejection and a better match to the signal's behavior than one-dimensional operations.

## 2.1 Preprocessor

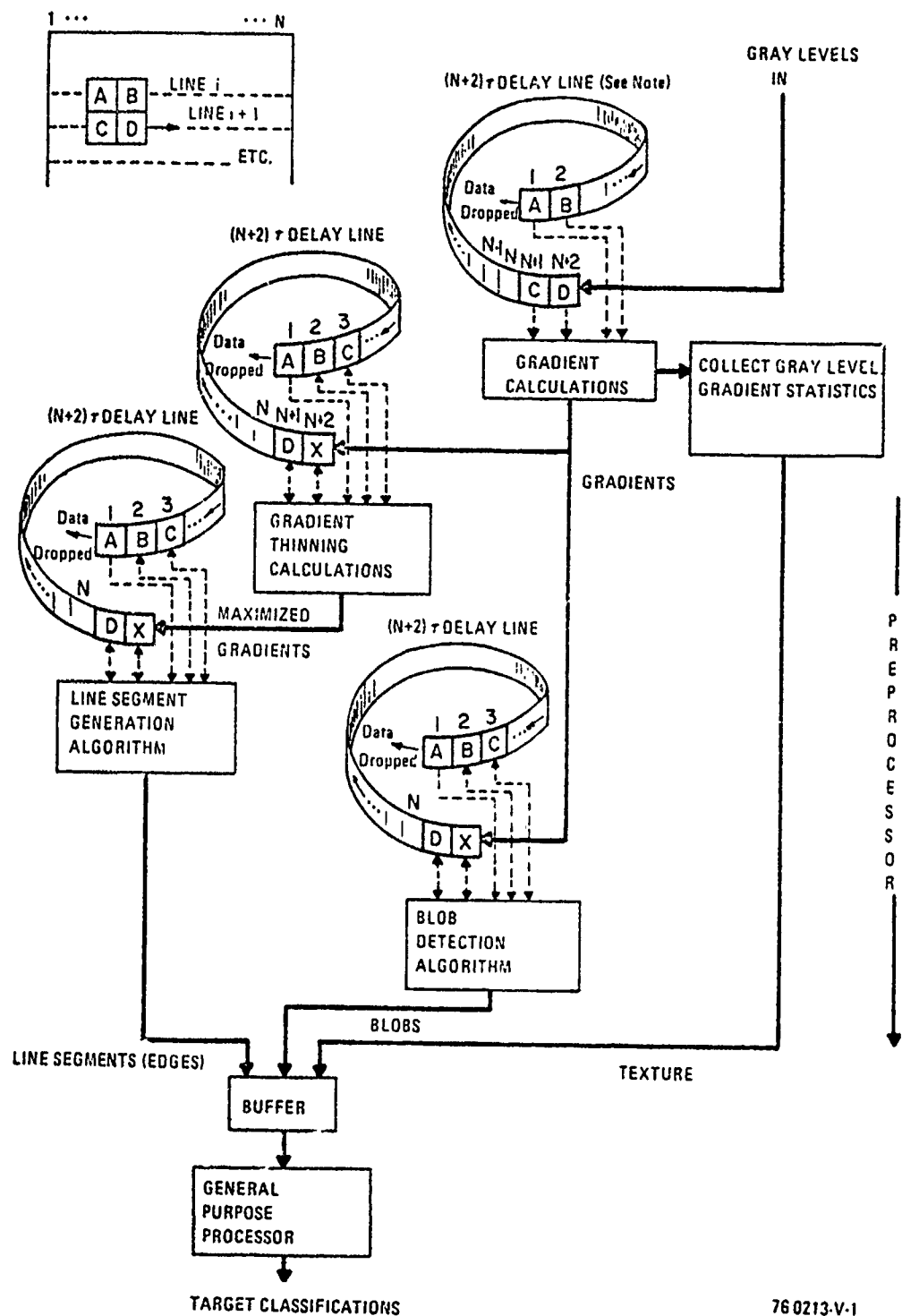
The function of the preprocessor is to extract from the gray level image the information required for generating recognition features. Three types of data are extracted. The primary data are the straight-line contours of gray-level gradient. Thus a line-drawing of the video image is generated. The second type of data are positional cues of gray-level closed objects (or "blobs"). The location of a blob generates a window within which recognition features are generated. The final set of data are statistical parameters computed during the preprocessing which may be used in texture classification.

Operation of the preprocessor is on a line-by-line basis with respect to the input image. Therefore, video data may be handled directly. Furthermore, storage requirements in the preprocessor are limited to single lines of data only.

### 2.1.1 Gradient Extraction

To generate the straight-line contours (subsets) of the image, it is necessary to first compute the two-dimensional gradient at each image point.

NOTE: DELAY LINE STORAGE IS  
EQUIVALENT TO RASTER SCAN OF  
2-D IMAGE



76 0213-V-1

Figure 2-3. Digital Image Processor

This is done as shown in Figure 2-4 with a four-pixel window scanning across the image in a raster format. The gradient amplitude and angle are approximated as shown. The gradient direction is quantized to 16 discrete directions, as depicted in the diagram. To suppress the areas of negligible gradient activity (containing no significant contour or edge information), a threshold is applied to the gradient amplitude. Figures 2-5(a) and 2-5(b) show a gray level image and its computed gradient. This is a FLIR image of a small truck. The gradient image has been thresholded and displays gradient direction, with the directions 10 through 16 coded with an overprinted slash /.

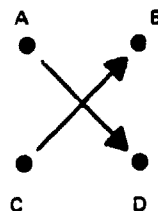
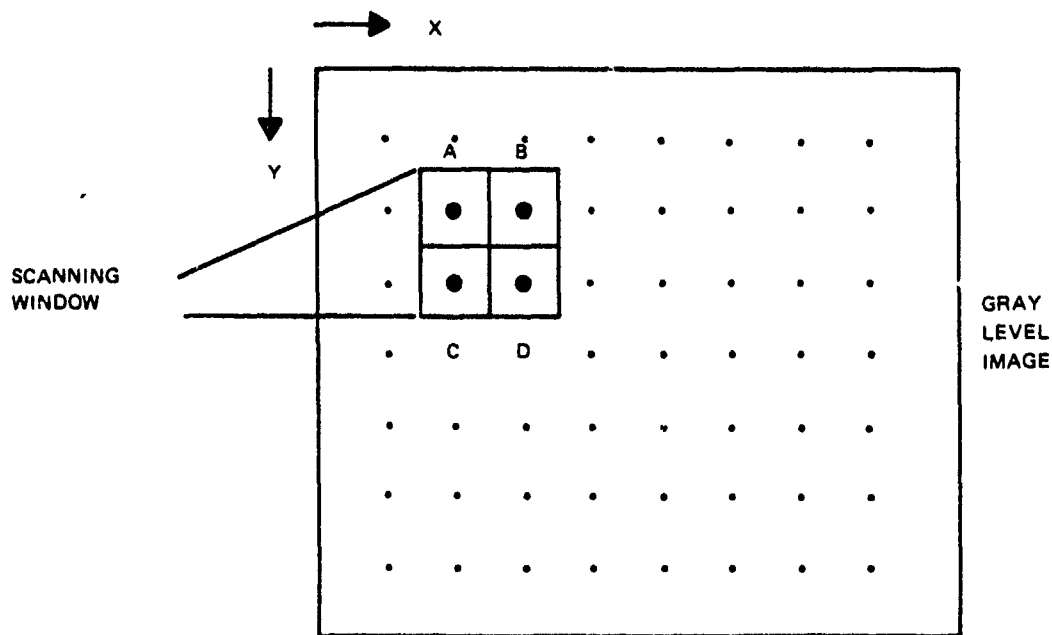
### 2.1.2 Gradient Maximizing

After gradient thresholding the edges are generally still too wide for subset generation. Therefore a gradient thinning operation is performed. The operation basically "skeletonizes" adjacent colinear gradient directions to the peak or maximum points.

The algorithm utilizes a raster scanning window containing a gradient cell "X" and 4 of its nearest neighbors. The scanning window is depicted at the top of Figure 2-6. The neighbors with colinear gradients are compared to "X". The largest gradient is then retained. This procedure is repeated sequentially for each gradient point in the image. An example of the maximized gradient is shown in Figure 2-5(c).

### 2.1.3 Subset Generation

Subset generation is accomplished by "growing" a line formed by adjacent parallel gradients. As before, a 5-cell scanning window is employed. The new data point is labeled cell "X". Its four neighbors are examined (sequentially: A, B, C, then D) to



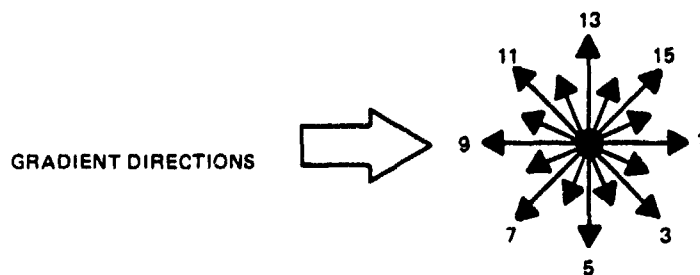
#### GRADIENT COMPONENTS

$$\Delta X = B - C$$

$$\Delta Y = D - A$$

$$\text{GRADIENT AMPLITUDE} = \text{MAX.} (\Delta X, \Delta Y) + \frac{1}{2} \text{MIN.} (\Delta X, \Delta Y)$$

$$\text{GRADIENT DIRECTION} = \left\lceil \tan^{-1} \left( \frac{\Delta Y}{\Delta X} \right) \right\rceil, \text{ QUANTIZED INTO 1 OF 16 DIRECTIONS FOR } 0 \rightarrow 2\pi$$



73-0985-V-5

Figure 2-4. Gradient Extraction Process

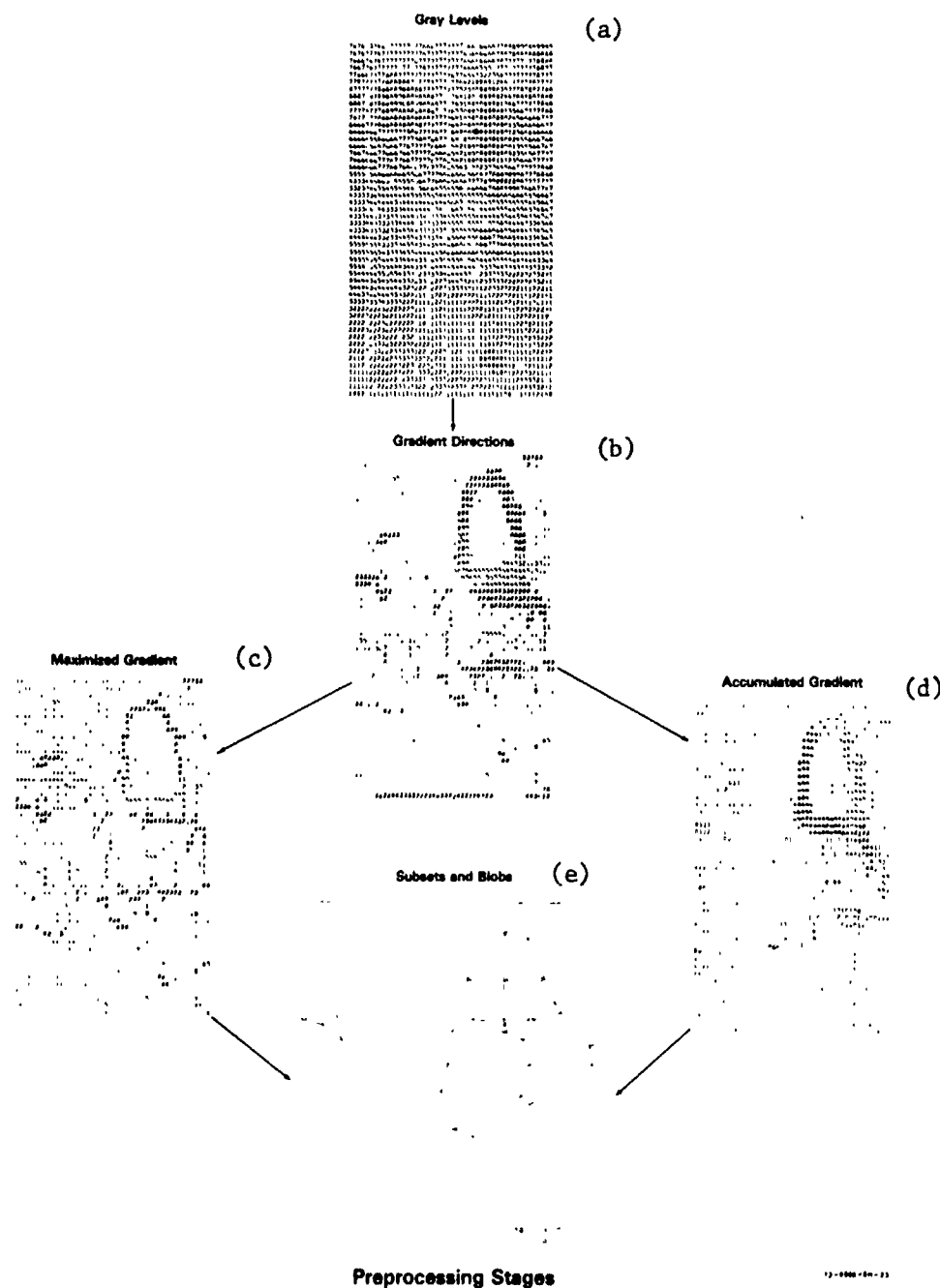


Figure 2-5. Steps in the Preprocessing of an Image

find those containing a parallel (within a tolerance) gradient direction. If one is found, then "X" is added as the next point in the line from the neighbor. Neighbors that are colinear to the gradient of "X" are excluded to prevent false lines from forming. The operational cycle of the subset generator is diagrammed in Figure 2-6. An example of the subsets derived from a gray level image is shown in Figure 2-5(e).

#### 2.1.4 Blob Detector

The blob detector detects the presence of a contiguous area of gray levels lighter (or darker) than its surrounding background. It operates independently of size, orientation, and position, and will detect all but sharp, concave shapes.

The operation of the blob detector is similar to that of the subset generator. The input data is the output of the gradient stage. Basically, the blob detector seeks to trace paths along contiguous, slowly changing gradient directions. Bookkeeping counters for each path being traced keep track of the gradient at the start of the path. When two paths from the same starting gradient join, a blob detection occurs. Additional bookkeeping counters measure the maximum and minimum X and Y excursions, providing a measure of the blob's size.

Figure 2-6 depicts the operational cycle of the blob detector. It uses the basic 5-cell window scanning the gradient image. Each of the 4 neighbors of the X-pixel is examined to determine if X should be added as the next point in a blob tracing path. Figure 2-5(d) displays the paths being traced out from the gradient image, Figure 2-5(b). The numbers

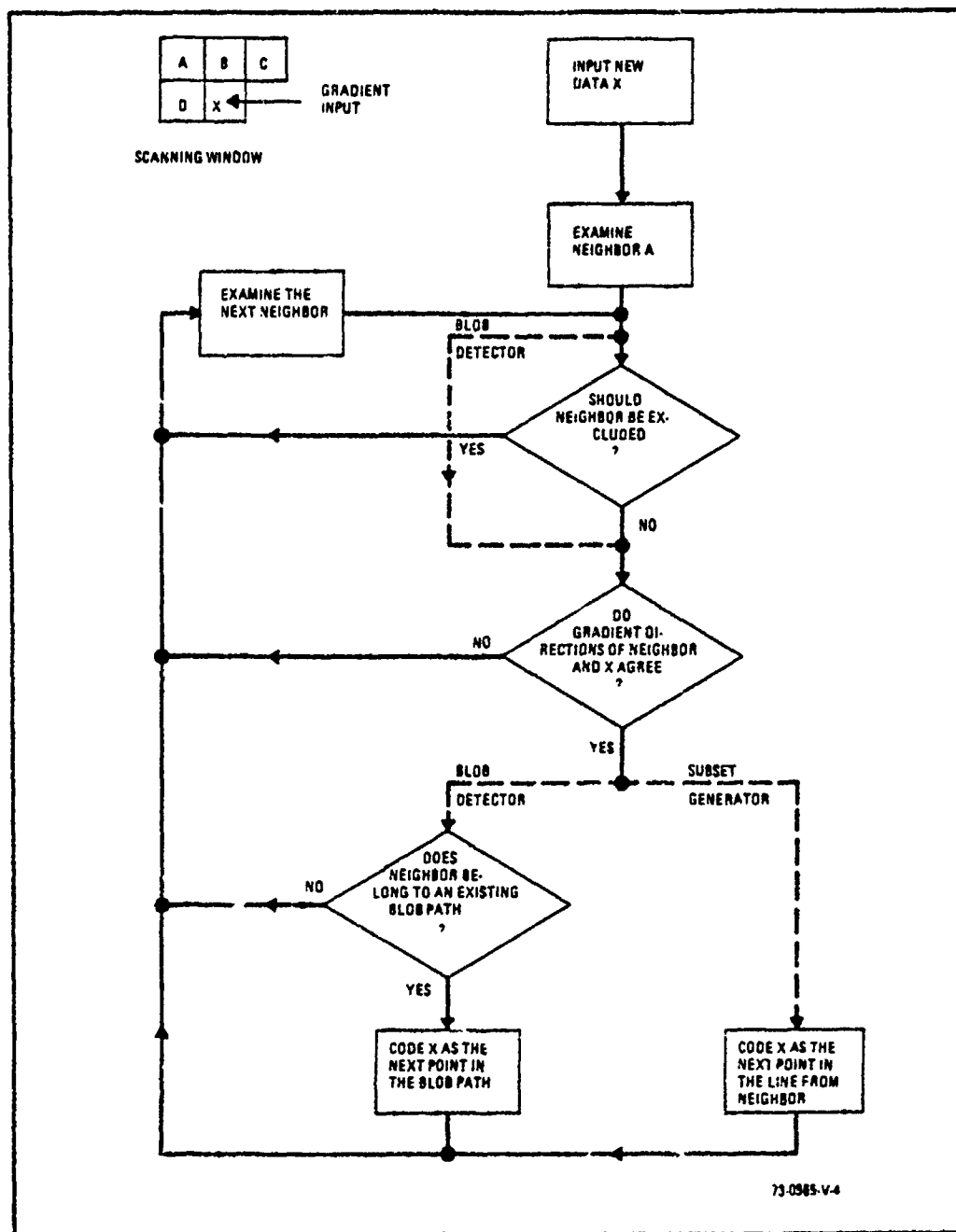


Figure 2-6. Block Diagram - Blob Detector and Subset Generator - Operational Cycle

printed out in the figure indicate the coded bits that keep track of the start of each path. Blob detection is coded as a pair of B's. The output of the blob detector consists of the blob polarity, center position, and horizontal and vertical dimensions. This data permits the object to be isolated for feature extraction.

#### 2.1.5 Texture Data

The third preprocessor function is the collection of statistical data for texture classification. The gray level image area is divided into windows of 30 x 30 pixels for statistical data collection. The average gray level and average gradient amplitude is computed. A limited histogram of the gradients is accumulated; i.e., the number of pixels with gradient amplitude equal to zero, one, two, and three. Also, two additional parameters are computed: (1) the number of pixels with gray level  $> a$ , and (2), the number of pixels with gray level  $< b$ . The subset generator provides two statistics: (1) the number of subsets per window, and (2), the number of "long" subsets.

This study, however, concentrated on the training and testing of the target recognition algorithms, not so much on texture analysis. The texture statistics were generated during the study, but were not classified or utilized.

#### 2.2 Final Processor

The final processing of the data is accomplished in a programmable processor (general-purpose computer). Its task is to generate the recognition features and perform the target decision logic. A block diagram of the final processor is shown in Figure 2-7.

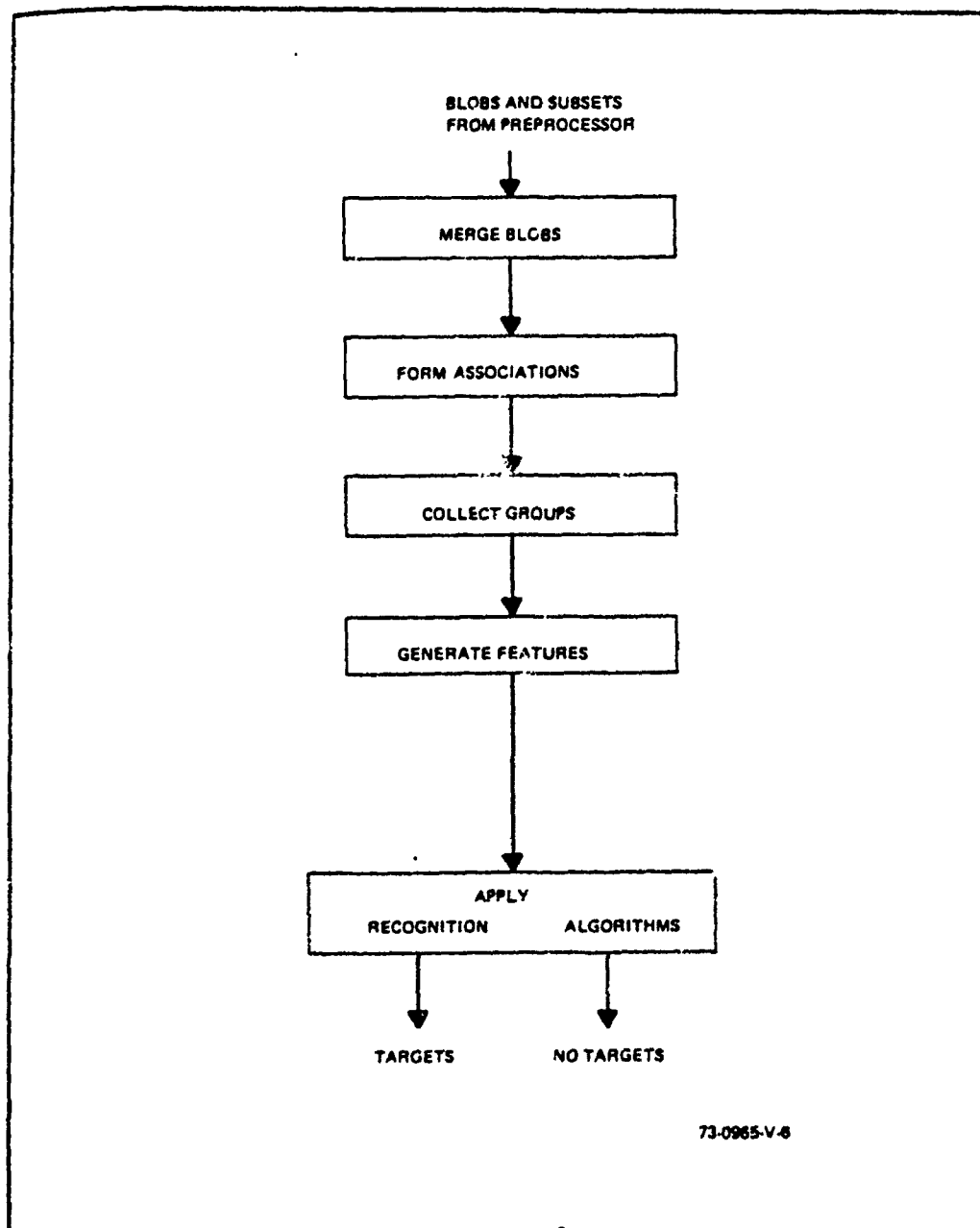


Figure 2-7. Final Processor

### 2.2.1 Blobs and Groups

To reduce storage and speed requirements and to reduce background interference, recognition features are not computed for the entire image. Instead, they are initially computed only within local rectangular areas whose positions are designated by the blob detector. Therefore, the blob detector in effect "cues" the processor to a local area containing a possible target. However, for those targets having complex shapes, such as aircraft, cues are also initiated by the presence of a "starter" subset. A starter subset is defined as one whose length exceeds a predetermined value (e.g.,  $l > 5$ ). For each starter subset within the image, a square area (or window) centered on the subset is also used as a positional cue for the processor. The blob and long subset windows are used to collect groups of subsets, as will be discussed later.

As seen from Figure 2-7, the first function performed by the processor is blob merging. Under certain conditions a single target can give rise to multiple (usually no more than two) blob detections that overlap. Therefore, the blob list in the preprocessor buffer stage is scanned for blobs with overlapping areas. Overlapping blobs are merged into a single new blob whose area will enclose the union of the original blob areas. See Figure 2-8.

Following blob merging a search is made for several different "associations". In general, an "association" means that an element (e.g., blob) is within a specified distance from another element. An association of long subsets with other long subsets is a significant association. These pairings may later be screened to detect the presence of roads. Also the association of blobs with

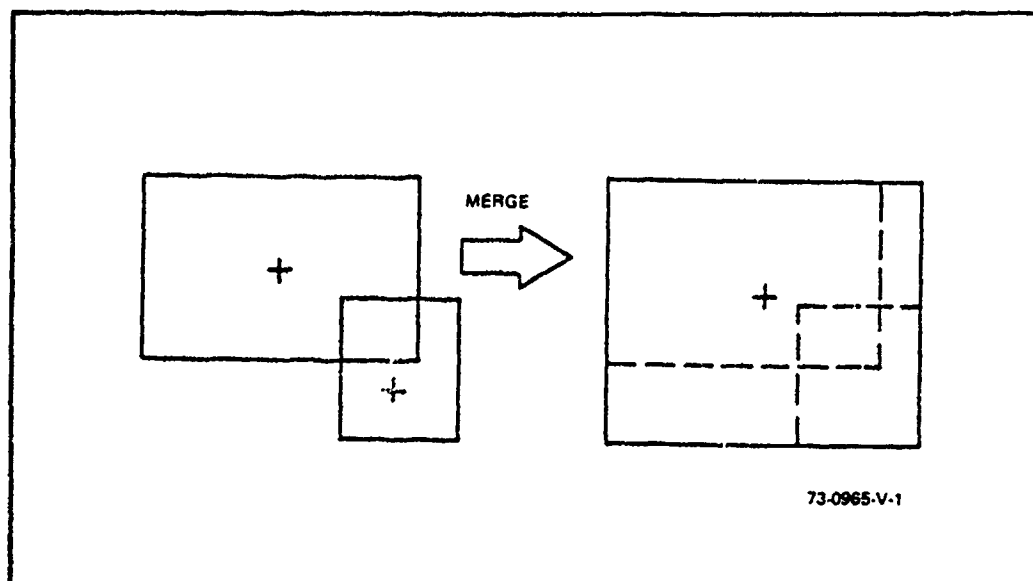
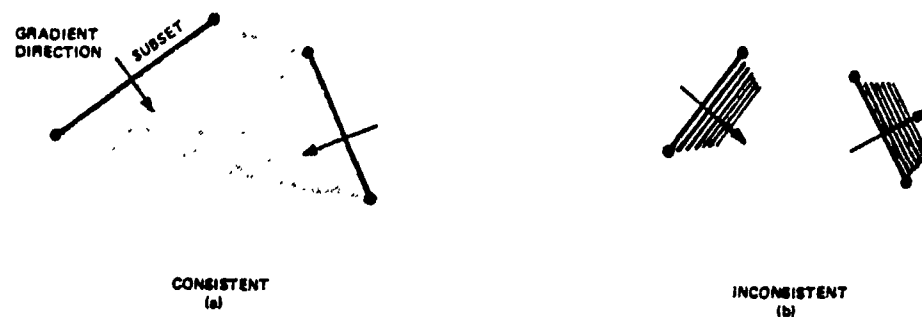


Figure 2-8. Blob Merging

long subsets is examined. Subsequently, these long subsets are prevented from being used to collect a subset group, since the blob is usually a more accurate cue.

When the associations have been made the process of group forming starts. Each blob or long subset defines a window. For each window, all subsets are screened by X-Y position. All the subsets falling within the window are defined as the group for that window.

Further screening of the groups is done to eliminate subsets not belonging to a candidate target area. It should be noted that the gradients of the subsets belonging to any dark (light) object point inwards (outwards), with few exceptions. See Figure 2-9(a).



73-0965-V-2

Figure 2-9. Significance of Polarities Between Subsets

Subset pairs with non-opposing (inconsistent) polarities, as in Figure 2-9(b), do not usually belong to an object, but are merely background clutter. Therefore, long subset groups are screened of any subsets with polarities inconsistent with the long subset defining the group. Blob groups are screened of any subsets with polarities inconsistent with the blob color, and relative to its center.

### 2.2.2 Feature Generation

The performance of a recognition system ultimately depends on the choice of measurements or features representing the target which are used by the decision logic. Because of its programmable nature, the final processor can be readily modified as regards both the target complement and their associated feature sets.

The training phase of this study resulted in the selection of 11 types of features to be calculated for each blob or long subset group (i.e., candidate object). These will now be described.

(1) Dimensions:

The vertical extent  $\Delta Y$  of an object is output from the blob detector or computed from the long subset group. The horizontal extent  $\Delta X$  is also computed.

(2) Aspect Ratio:

The aspect ratio is defined as  $\Delta X / \Delta Y$ .

(3) Number for Further Processing - N.F.P.:




As previously stated, the subsets in each group are screened for polarity. In addition, the remaining subsets are designated as belonging to the top or to the bottom of the group. The designation is based upon the orientation and polarity direction of the subset. This sorting effectively separates the object into a bottom half and a top half. In the process, if any subset's midpoint physically occurs in the opposite half, it is thrown away. The number of subsets remaining at this point is called the N.F.P. count.

(4) Final Active Quadrant Count - NFACT

The subsets are also sorted into a right side or a left side based on angle and polarity. At this point, each subset has been assigned to one of four quadrants. The number of quadrants that have at least one subset is the NFACT count.

(5) Length Residue  $\Delta R$ :

This is a feature useful for separating triangular objects from rectangular objects. It is an approximation of the total length of

non-parallel subsets within a group. Its dimension is in pixels. A triangle  has a positive value, a rectangle  has zero value, and a triangle  has a negative value. It is computed as follows:

For each quadrant  $K = 1, 2, 3, 4$  compute

$$R_K = N_K \frac{\sum S \cdot d \cdot \frac{\alpha}{45}}{\sum d}$$

where  $N_K$  = number of subsets in quadrant  $K$

$S$  = the subset's length

$\alpha = \min. \begin{cases} \text{off-vertical orientation of the subset} \\ \text{off-horizontal orientation of the subset} \end{cases}$

$d$  = distance from center of group.

$$\text{Then } \Delta R = \left( R_{\text{left top}} + R_{\text{right top}} \right) - \left( R_{\text{left bot.}} + R_{\text{right bot.}} \right)$$

(6) Closure:

Closure is defined as  $\frac{\sum S}{P}$

where  $S$  = each subset's length

$$P = 2 \cdot (\Delta X + \Delta Y)$$

(7) KHOLE:

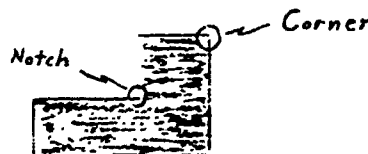
Many APC targets display a black "hole" from the rear viewing angle. This feature searches for this property. If a subset of the correct angle and polarity is found in the top half of an object, such as to be the top part of the "hole", then KHOLE = 1. Otherwise, it is 0.

(8) LONTOP:

Many of the tanks at long range displayed a rather long, somewhat horizontal top. (Apparently the turret was not very hot.) So if a long, nearly horizontal subset forms the very top of the object, LONTOP = 1. Otherwise, it is 0.

(9) Corners and Notches:

The top half and bottom half of the object are searched separately for the presence of an outside corner or an inside corner (notch). See the drawing below.



Those nearest  $90^\circ$  are printed out. We thus have four possibilities:

$C_{top}$ ,  $C_{bot}$ ,  $N_{top}$ ,  $N_{bot}$ .

(10) Peak:



To discriminate tall column-like tops (or bottoms) from low broad or flat tops (or bottoms), the peak calculation is made for the subsets in the top and bottom halves, separately. It is computed as:

$$\text{Peak} = \frac{\sum \beta \cdot S/2}{\sum S/2} \times 100$$

$$\text{where } \beta = 2 \left( \frac{\theta}{90^\circ} \right) - 1$$

$\theta$  = subset's horizontal angle

$S$  = subset length

Thus the shape  has negative Peak, while  has positive Peak values. One other parameter is computed:  $\Delta \text{Peak} = \text{Peak}_{top} - \text{Peak}_{bot}$ .

(11) Sym:

To measure the symmetry of the top (or bottom) of an object, another calculation is made. For the top and the bottom subsets, compute:

$$\text{Sym} = \frac{\sum Z \cdot (X - X_0)^3}{\left(\frac{\Delta X}{2}\right)^3 \sum Z} \times 100$$


$$\text{where } Z = \frac{S}{2} \cdot \left( \frac{90^\circ - \theta}{90^\circ} \right) + \frac{\theta}{90^\circ}$$

$S$  and  $\theta$ , as before,

$X_0$  = midpoint of the object

$X$  = the subset's midpoint.

Asymmetric tops, such as  will have a large Sym magnitude.

Symmetric tops  will have zero values. Two additional parameters are computed:

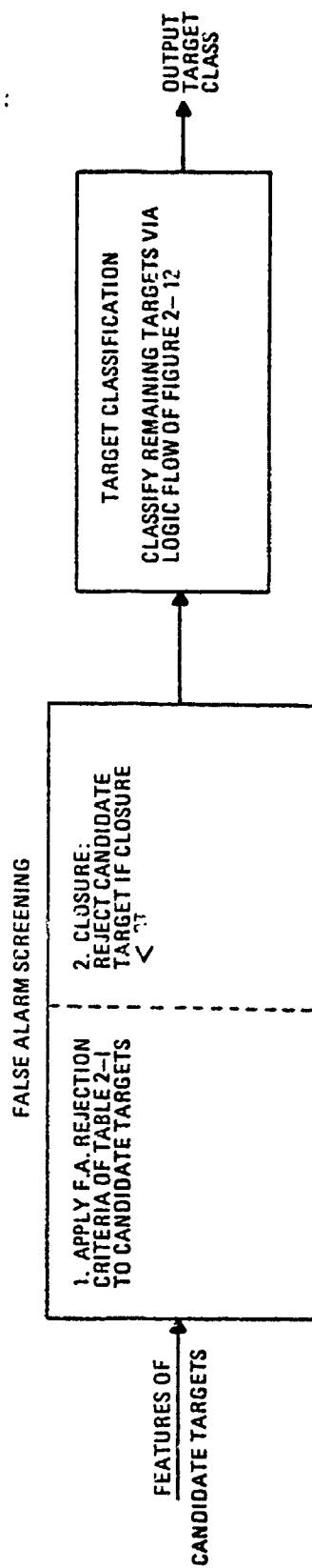
$$\overline{S_{ym}} = \frac{1}{2} \left\{ |S_{ym_{top}}| + |S_{ym_{bot}}| \right\}$$

and

$$\Delta S_{ym} = \left| \left\{ |S_{ym_{top}}| - |S_{ym_{bot}}| \right\} \right|$$

### 2.2.3 Recognition Algorithm

As indicated earlier in Figure 2-7, the final block in the processor is the recognition algorithm. A block diagram is shown in Figure 2-10. The features for each blob or long subset group have now been computed. The first process indicated in the figure is the screening out of false alarms, or non-targets. To that end, two stages are employed. The first stage uses the False Alarm Rejection Criteria of Table 2-I. A failure in any of the rules, rejects the group as a false alarm. The second stage is a minimum acceptable value for the feature Closure, as shown in Figure 2-10.



76-0564-V-3

Figure 2-10 Recognition Algorithm - Block Diagram

TABLE 2-I

FALSE ALARM REJECTION CRITERIA

REJECT A CANDIDATE TARGET AS A FALSE ALARM  
IF IT FAILS ANY FOLLOWING TEST:

- 1)  $NFACT \geq 3$  (for subset groups, only)
- 2)  $NFP \geq 3$  (for subset groups, only)
- 3) C-N count
 

{	$\geq 1$ for $.37 \leq \text{Closure} \leq .6$ $\geq 2$ for $.6 < \text{Closure} \leq .7$ $\geq 3$ for $.7 < \text{Closure} \leq .8$
---	--

  
(for subset groups, only)
- 4)  $\text{Closure} \leq .8$  (for subset groups, only)
- 5)  $|\Delta R| > .19$
- 6)  $.4 \leq \text{Aspect Ratio} \leq 4.$  (for blob groups)  
 $.4 \leq \text{Aspect Ratio} \leq 3.$  (for subset groups)
- 7)  $\text{Peak}_{\text{Bot}} \leq 30.$
- 8)  $\left| \text{Sym}_{\text{Top}} \right| < 900.$   
       or Bot
- 9)  $3 \leq \Delta X \leq 26$   
        $3 \leq \Delta Y \leq 26$

All groups (or objects, at this point) remaining are considered targets and will be classified into one of the three target classes. A total of 13 features are used to classify the targets. Normally, though, only the first 11 are used in classification; the remaining 2 are added for tie-breaking cases. Decision boundaries in the 11-dimensional\* feature space were established during the training phase, as described in Section 3.4. To achieve an early estimate of performance, a simplified decision space was utilized. As shown in Figure 2-11, nine features are used in a pair-wise manner to yield 15 separate classification regions. The 10th and 11th features (KHOLE and LONTOP) provide 2 additional classification regions.

The first step in classifying a target is to determine the region (or regions) that contains the target's feature pattern, to provide a tentative class decision(s). As shown in Figure 2-12, the next step is to take a vote of the tentative decisions. Note that a NON-TRUCK region provides one TANK vote and one APC vote. Similarly, NON-APC and NON-TANK provide 2-vote tentative decisions.

If there is no majority, special tie-breaking rules are employed. These are tabulated in Table 2-II, and utilized as shown in Figure 2-12. The final classification decision, as well as the target's coordinates are the output data.

---

\*However, one of the features,  $\Delta$ Peak, is correlated with two other features, Peak<sub>TOP</sub>, and Peak<sub>BOT</sub>.

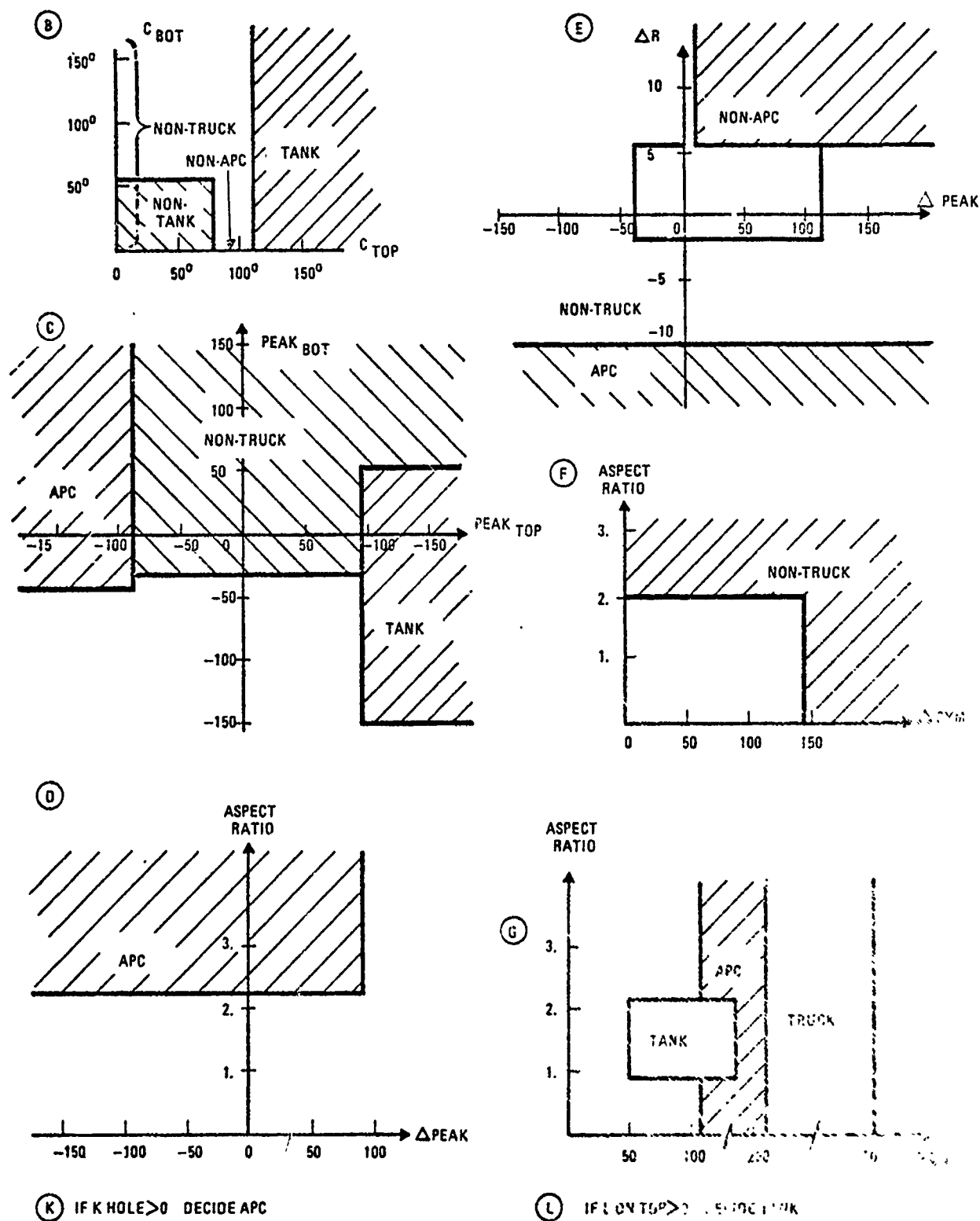
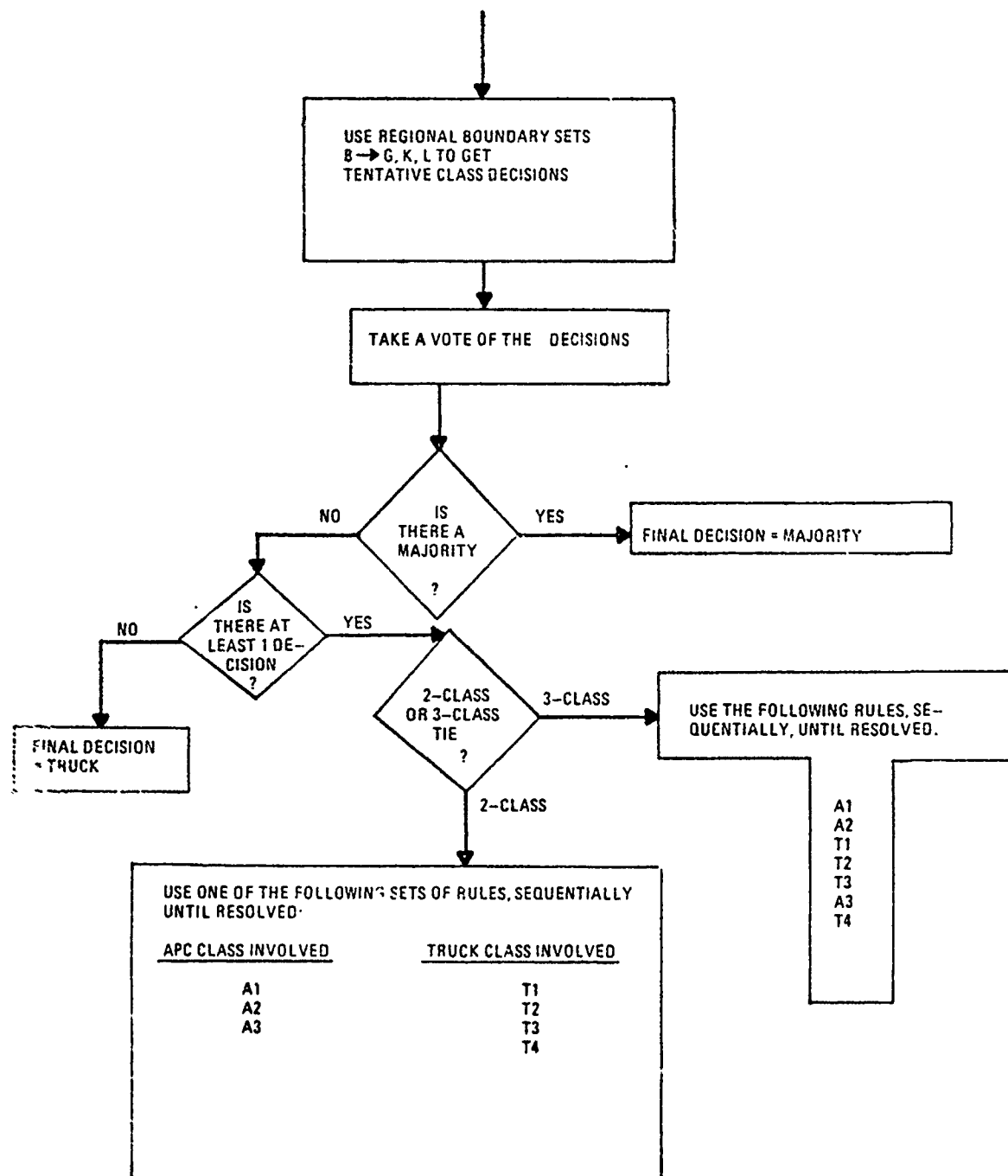


Figure 2-11 Regional Boundary Sets B -- L, K, L



76-0564-V-4

Figure 2-12 Classification Logic Flow

TABLE 2-II

TIE-BREAKING RULES

## SET A

- |    |    |   |                |
|----|----|---|----------------|
| 1: | If | $\Delta \text{Peak} < -50$                            | decide APC     |
| 2: | If | $\text{Peak}_{\text{Top}} < \text{Peak}_{\text{Bot}}$ | decide non-APC |
| 3: | If | $\text{N.F.P.} \geq 9$                                | decide APC     |

## SET T

- |    |    |                                 |                  |
|----|----|---------------------------------|------------------|
| 1: | If | $\Delta \text{Peak} \geq 110$   | decide non-Truck |
| 2: | If | $\text{Peak}_{\text{Top}} > 10$ | decide non-Truck |
| 3: | If | there exists $N_{\text{Bot}}$   | decide non-Truck |
| 4: | If | $\Delta R > 0$                  | decide Truck     |

### 3.0 TEST PROGRAM USING NVL IMAGERY

The statistical test program involved the collection and preparation of the data base, and the use of these images to "train" and "test" the image processing system by computer simulation. These steps will be described in this section.

#### 3.1 The Data Base

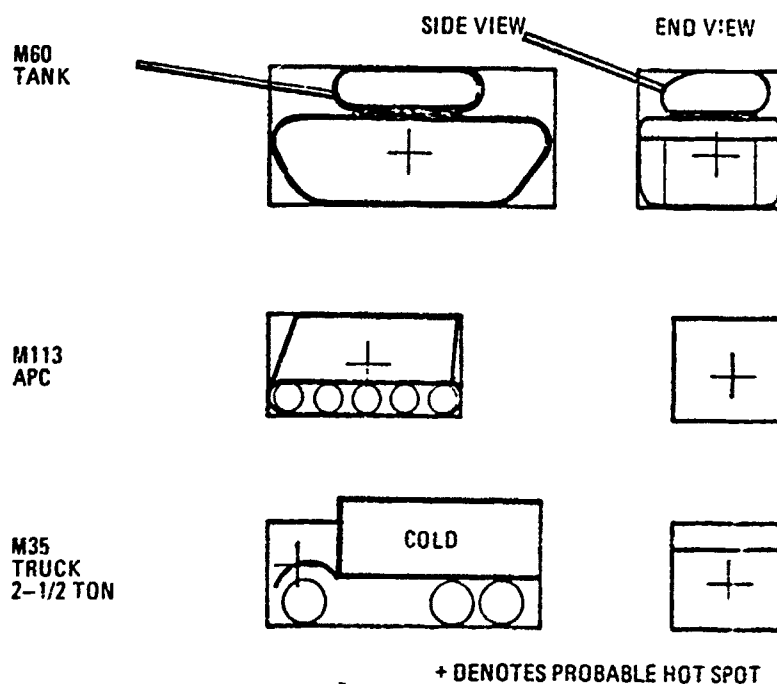
The data base was supplied by NVL. It consists of 14 magnetic digital tapes of FLIR data, as listed in Table 3-I. The images had been digitized from a TV-format FLIR system via video tape recordings. Each digitized image is 800 x 1024 picture elements (pixels) of 8-bit gray level data. Each digital tape file contains a separate image. Ground truth and 35-mm. film transparencies were also supplied for the images.

The imagery contains target and a few non-target scenes. The targets are: an M60A tank, M113 APC, and a 2½ ton truck (probably M35 type). Rough sketches of these targets are shown in Figure 3-1. Dimensional information is also given. Probable IR "hot" spots are located by the "+" signs. A study of the film strips that were provided shows that frequently at longer ranges, the turret of the tank is not visible. The truck has a "cold" area in the rear, noticeable at close range. The APC also has a distinct characteristic, at close range and rear view. It has a noticeable black hole in the middle, where the door is located. However, at moderate and longer ranges the targets are difficult to visually recognize. (This will be discussed further in a later section.)

TABLE 3-I LIST OF DIGITAL TAPES AVAILABLE

<u>TAPE</u>	<u>NUMBER OF FILES</u>
A	10
B	10
C	10
D	10
E	10
F	10
G-H	15
H-I	15
I-J	15
J-K	15
L	10
M-N	15
N-O	10
P	> 12

	HEIGHT (M)	LENGTH (M)	WIDTH (M)
TANK M60 A	3.26	6.95	3.63
APC M113	2.2	4.68	2.69
TRUCK M35	2.54	6.71	2.44



76-0564-V-6

Figure 3-1 Sketch of Targets

The film strips also reveal that many of the images are seriously degraded in quality. Vertical stripes and ripple are present on the left side of images. Herringbone and Moiré patterns, and ripple appear sporadically over the field-of-view of many images, and horizontal black and white streaks occur occasionally. Additionally, the resolution appears to be much lower than the pixel spacing. These distortions will be considered later.

### 3.2 Preparation of Imagery

As the digital tapes arrived from NVL, they were copied to provide "working" tapes more compatible with the particular tape drives at Westinghouse. Copying the tapes was often a difficult task; errors were frequently encountered. On the average, two attempts per tape had to be made. It was also discovered that Tape P contained completely unknown data. It was therefore dropped from the data base.

The second step towards preparing the data base was to generate a set of sub-images or "windows". The existing simulation of the image processing system uses images of size 50 pixels by 50 pixels. This heretofore provided a more than adequate area to include any target of interest, plus some background clutter. It is also fast-running in the simulation, keeping computer time costs at a minimum. To hold computer costs down and stay compatible with the existing software, the same size format was used for this study.

A 50 x 50 pixel window was created for each target in each image. Using the ground truth information and film strips that were supplied, the coordinates of 50 x 50 size areas containing a target were tabulated. Using a computer subroutine, the gray levels of these areas were "lifted" from the

digital tapes and copied onto another magnetic tape, as separate files. Figure 3-2 shows an example of a window containing an APC lifted (or extracted) from file 2 on magnetic tape I-J (ground truth image L-10).

Some of the targets in the data base are very large (e.g., > 100 pixels in length). To fit them within the 50 x 50 windows, areas containing larger targets were digitally shrunk to approximately 15-20 pixels in target length. The shrinking was done by averaging a neighborhood and using that value as a single new pixel. A 2:1 shrink, for example, averages a neighborhood of 2 x 2 pixels to obtain a gray level. The next gray level comes from the next adjacent 2 x 2 neighborhood. As a consequence, high frequency noise is reduced and resolution is reduced. However, the resolution loss was considered non-detrimental for two reasons. First, the FLIR sensor data had been oversampled in deriving the digital version of the images. Secondly, the present target recognition system is oriented towards operation with longer range targets - thus small size (10-30 pixels, e.g.) and lower resolution-on-target. The result, then, of those images that were shrunk while being extracted is a smaller, somewhat smoothed, version of the original.

At the same time that the windows were being extracted from the data tapes, the lowest 2 bits of gray level data were dropped. The digital image processing system requires only 5 or 6 bits of data, and it was estimated that the image data provided did not contain any significant target or scene information in the lowest 2 bits. So only the upper 6 bits were retained.

After the 50 x 50 windows were written on magnetic tape, they were photographically played back for visual inspection and verification. The playback photos revealed that Tapes I-J and J-K did not coincide with their expected film strip images. Tapes I-J and J-K had to be reformatted and

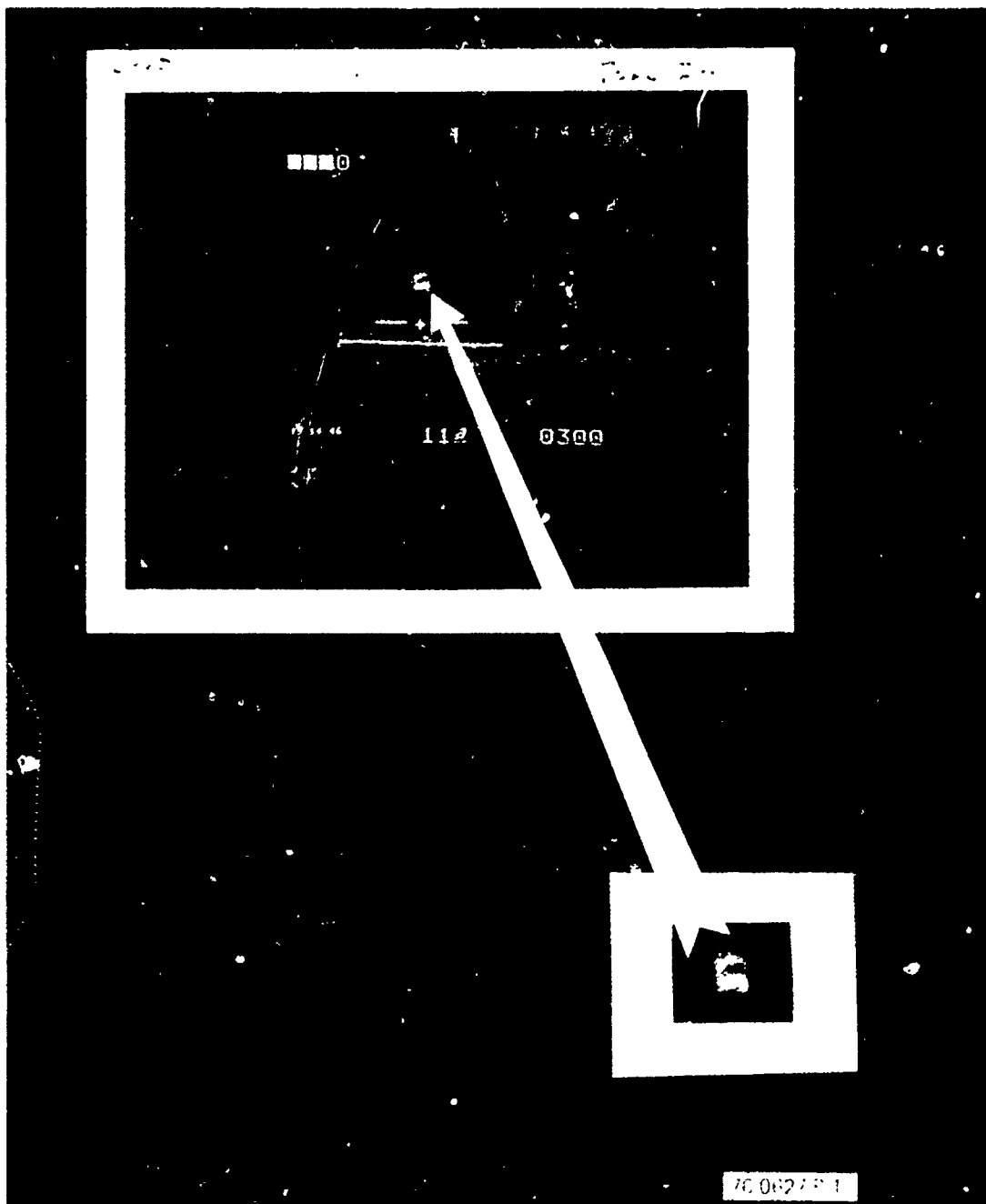


Figure 3-2. A 50 x 50 Window Extracted from an Image

recopied to make tapes compatible with the photo playback system. Playbacks of these tapes showed that Tape I-J contained images L-9, L-10, J-7 through J-10, and L-1 through L-9, in that order. Tape J-K contained, consecutively, images J-1 through J-10, and K-1 through K-5 (instead of J-6 to J-10, K-1 to K-10 as expected).

A total of 1005 windows were extracted. Approximately 240 contain targets (some are unknown in ground truth); the remaining windows contain no targets and are used for false alarm testing. Table 3-II lists all of the windows and targets, along with some diagnostic and ground truth data. The last page of Table 3-II lists the sources of most of the non-target windows.\* The "MULX" and "MULY" columns indicate that a whole set of adjacent 50 x 50 windows were extracted from one image. For example, the last entry indicates that 320 windows (20 across by 16 down) were listed from image D-9 and were labeled window number 686 through 1005. Figure 3-3 shows photographic playbacks of all 1005 windows comprising the data base. It should be explained that the windows that appear to be all white are playbacks of windows extracted from images with reversed (negative) polarity. Upon extraction, these gray levels were complimented for polarity correction, however no d.c. adjustment was made since the preprocessor only uses the gradient information. Unfortunately, this sometimes caused white saturation during playback (with the brightness and contrast set up for normal polarity windows).

The statistical test requires separate sets of training and test images. Therefore the windows containing targets (as verified by the playbacks) were split about equally between training and test. An attempt was made to alternate successive images of a target between training and test, so that the two sets would contain roughly similar aspect angles, ranges, etc. for each target.

\*Some of the images did contain targets and had already been extracted, so those 50 x 50's here that contained such areas were excluded from any further usage.

TABLE 3-II LIST OF DATA WINDOWS

PICT. NO.	COMMENTS OR DESCRIPTION	L=	POL	SHK
1	TAPE A PIX 1 TANK/S	L=80	U	6
2	TAPE A PIX 2 TANK/S	L=80	U	6
3	TAPE A PIX 3 TANK/S	L=80	U	4
4	TAPE A PIX 3 2 1/2 TON/S	L=70	U	4
5	TAPE A PIX 4 TANK/S	L=80	U	5
6	TAPE A PIX 4 2 1/2 TON/S	L=80	U	5
7	TAPE A PIX 5 FALSE ALARM	L=70	U	4
8	TAPE A PIX 6 TANK/S	L=80	U	4
9	TAPE A PIX 6 2 1/2 TON/S	L=70	U	4
10	TAPE A PIX 7 FALSE ALARM	L=80	U	4
11	TAPE A PIX 8 TANK/S	L=80	U	5
12	TAPE A PIX 9 TANK/S	L=80	U	5
13	TAPE A PIX 9 2 1/2 TON/S	L=70	U	3
14	TAPE A PIX 10 TANK/S	L=100	U	6

PICT. NO.	COMMENTS OR DESCRIPTION	L=	POL	SHK
15	TAPE B PIX 1 TANK/S	L=50	0	4
16	TAPE B PIX 1 FALSE ALARM	L=40	0	4
17	TAPE B PIX 2 TANK/S	L=80	0	6
18	TAPE B PIX 2 FALSE ALARM	L=40	0	4
19	TAPE B PIX 3 FALSE ALARM	L=80	0	6
20	TAPE B PIX 3 TANK/S	L=30	0	2
21	TAPE B PIX 4 TANK/S	L=50	0	4
22	TAPE B PIX 5 TANK/S	L=80	0	5
23	TAPE B PIX 5 FALSE ALARM	L=80	0	7
24	TAPE B PIX 5 FALSE ALARM	L=80	0	5
25	TAPE B PIX 6 TANK/S	L=80 70-80	0	6
26	TAPE B PIX 7 TANK/S	L=80	0	6
27	TAPE B PIX 8 2 1/2 TON E	L=80	0	5
28	TAPE B PIX 9 FALSE ALARM	L=45 NEGATIVE	0	3
29	TAPE B PIX 9 FALSE ALARM	L=80 NEGATIVE	0	6
30	TAPE B PIX 10 FALSE ALARM	L=80 NEGATIVE	0	6

PICT. NO.	COMMENTS OR DESCRIPTION	L=	POL	SHK
31	TAPE C PIX 1 APC/S	L=50	0	3
32	TAPE C PIX 1 TANK/E	L=80	0	6
33	TAPE C PIX 2 APC/S	L=80	0	5
34	TAPE C PIX 2 TANK/E	L=80	0	4
35	TAPE C PIX 2 2 1/2 TON/E	L=80	0	5
36	TAPE C PIX 3 APC/S	L=50 NEGATIVE	1	5
37	TAPE C PIX 3 TANK/E	L=70 NEGATIVE	1	5
38	TAPE C PIX 3 2 1/2 TON/E	L=80 NEGATIVE	1	4
39	TAPE C PIX 4 APC/S	L=80	0	4
40	TAPE C PIX 4 TANK/E	L=80	0	6
41	TAPE C PIX 4 2 1/2 TON/E	L=80	0	6
42	TAPE C PIX 5 RETICLE	L=20	0	4
43	TAPE C PIX 6 2 1/2 TON/E	L=20	0	2
44	TAPE C PIX 6 TANK/S	L=80	0	5
45	TAPE C PIX 7 2 1/2 TON	L=30	0	3
46	TAPE C PIX 7 2 1/2 TON	L=20	0	2
47	TAPE C PIX 8 TANK/S	L=50	0	4
48	TAPE C PIX 9 TANK	L=50	0	4
49	TAPE C PIX 9 APC	L=25	0	2
50	TAPE C PIX 10 TANK	L=40	0	4
51	TAPE C PIX 10 APC	L=25	0	2

PICT. NO.	COMMENTS OR DESCRIPTION	L=	POL	SHK
52	TAPE D PIX 1 2 1/2 TON	L=25	0	2
53	TAPE D PIX 1 TANK/S	L=45	0	4
54	TAPE D PIX 2 2 1/2 TON	L=20	0	1
55	TAPE D PIX 2 TANK	L=20	0	2
56	TAPE D PIX 2 APC	L=20	0	1
57	TAPE D PIX 3 2 1/2 TON	L=15	0	1
58	TAPE D PIX 3 TANK	L=25	0	2
59	TAPE D PIX 3 APC	L=20	0	1
60	TAPE D PIX 4 2 1/2 TON	L=20	0	1
61	TAPE D PIX 4 TANK	L=25	0	2
62	TAPE D PIX 4 APC	L=20	0	1
63	TAPE D PIX 5 2 1/2 TON	L=20	0	2
64	TAPE D PIX 5 TANK	L=25	0	2
65	TAPE D PIX 5 APC	L=15	0	1
66	TAPE D PIX 6 NO TARGET	L=20 TWO FA	0	2
67	TAPE D PIX 7 APC/E	L=40	0	3
68	TAPE D PIX 8 APC/E	L=50	0	3
69	TAPE D PIX 8 TANK/S	L=100	0	8
70	TAPE D PIX 9 NO TARGET	L=40	0	4
71	TAPE D PIX 10 FALSE ALARM	L=40	0	4
72	TAPE D PIX 10 FALSE ALARM	L=20	0	2
73	TAPE D PIX 10 TANK	L=20	0	2

TABLE 3-II LIST OF DATA WINDOWS

PICT.	NO.	COMMENTS OR DESCRIPTION	POL	SHK
74	TAPE E PIX 1	2 1/2 TON/E L=80	0	3
75	TAPE E PIX 2	APC/E L=50	0	3
76	TAPE E PIX 2	TANK/3/4 L=100	0	7
77	TAPE E PIX 3	TANK/3/4 L=90	0	7
78	TAPE E PIX 4	FALSE ALARM L=30	0	3
79	TAPE E PIX 4	APC/E L=40	0	3
80	TAPE E PIX 5	APC/E L=40	0	3
81	TAPE E PIX 5	TANK/3/4 L=90	0	6
82	TAPE E PIX 6	APC/E L=40	0	2
83	TAPE E PIX 6	TANK/3/4 L=80	0	6
84	TAPE E PIX 7	2 1/2 TON/E L=60	0	5
85	TAPE E PIX 8	APC/E L=50	0	3
86	TAPE E PIX 8	TANK 3/4 L=90	0	7
87	TAPE E PIX 9	FALSE ALARM L=10	0	1
88	TAPE E PIX 9	FALSE ALARM L=60	0	5
89	TAPE E PIX 10	FALSE ALARM L=20	0	N
90	TAPE E PIX 10	FALSE ALARM L=20	0	N
91	TAPE E PIX 10	TANK L=30	0	N
92	TAPE E PIX 10	APC L=20	0	N
93	TAPE E PIX 10	FALSE ALARM L=60	0	N
94	TAPE E PIX 10	FALSE ALARM L=50	0	N

PICT.	NO.	COMMENTS OR DESCRIPTION	POL	SHK
95	TAPE F PIX 1	APC L=20	0	2
96	TAPE F PIX 1	TANK/S L=30	0	2
97	TAPE F PIX 1	2 1/2 L=15	0	1
98	TAPE F PIX 2	APC L=10	0	1
99	TAPE F PIX 2	TANK/S L=30	0	2
100	TAPE F PIX 2	2 1/2 L=10	0	2
101	TAPE F PIX 3	APC L=20	0	2
102	TAPE F PIX 3	TANK/S L=25	0	2
103	TAPE F PIX 3	2 1/2 L=10	0	1
104	TAPE F PIX 4	APC L=20	0	2
105	TAPE F PIX 4	TANK/S L=25	0	2
106	TAPE F PIX 4	2 1/2 L=10	0	1
107	TAPE F PIX 5	APC L=15	0	1
108	TAPE F PIX 5	TANK/S L=20	0	2
109	TAPE F PIX 5	2 1/2 L=10	0	1
110	TAPE F PIX 5	FALSE ALARM L=10	0	1
111	TAPE F PIX 6	2 1/2 L=15	0	1
112	TAPE F PIX 6	TANK L=5	0	1
113	TAPE F PIX 6	APC L=20	0	1
114	TAPE F PIX 7	2 1/2 L=15	0	1
115	TAPE F PIX 7	TANK L=5	0	1
116	TAPE F PIX 7	APC L=20	0	1
117	TAPE F PIX 8	2 1/2 L=15	0	1
118	TAPE F PIX 8	TANK L=5	0	1
119	TAPE F PIX 8	APC L=20	0	1
120	TAPE F PIX 9	2 1/2 L=15	0	1
121	TAPE F PIX 9	TANK L=5	0	1
122	TAPE F PIX 9	APC L=20	0	2
123	TAPE F PIX 9	FALSE ALARM L=40	0	2
124	TAPE F PIX 9	FALSE ALARM L=70	0	4
125	TAPE F PIX 10	APC/E L=60	1	4

PICT.	NO.	COMMENTS OR DESCRIPTION	POL	SHK
126	TAPE G-H PIX 1	APC/E L=40	1	3
127	TAPE G-H PIX 1	TANK 3/4 L=75	0	5
128	TAPE G-H PIX 2	TANK/S L=17	0	1
129	TAPE G-H PIX 2	F.A. L=5	0	1
130	TAPE G-H PIX 3	TANK/S L=17	0	1
131	TAPE G-H PIX 3	F.A. L=5	0	2
132	TAPE G-H PIX 4	TANK/S L=20	0	1
133	TAPE G-H PIX 4	F.A. L=8	0	1
134	TAPE G-H PIX 5	2 L=10	1	1
135	TAPE G-H PIX 5	7 L=5	1	2
136	TAPE G-H PIX 5	TANK L=20	1	1
137	TAPE G-H PIX 6	7 L=10	1	1
138	TAPE G-H PIX 6	7 L=5	1	2
139	TAPE G-H PIX 6	TANK L=20	1	1
140	TAPE G-H PIX 7	7 L=15	1	1
141	TAPE G-H PIX 7	7 L=5	1	1
142	TAPE G-H PIX 7	TANK L=20	1	1
143	TAPE G-H PIX 8	7 L=10	0	1
144	TAPE G-H PIX 8	7 L=5	0	1
145	TAPE G-H PIX 8	TANK L=20	0	1
146	TAPE G-H PIX 9	7 L=10	0	1
147	TAPE G-H PIX 9	7 L=5	0	1
148	TAPE G-H PIX 9	TANK L=20	0	1
149	TAPE G-H PIX 10	F.A. L=40	0	4
150	TAPE G-H PIX 10	F.A. L=20	0	2
151	TAPE G-H PIX 11	TRUCK 3/4 L=100	1	8
152	TAPE G-H PIX 12	TRUCK 3/4 L=100	1	8
153	TAPE G-H PIX 13	APC/E L=40	1	3
154	TAPE G-H PIX 13	TANK 3/4 L=70	1	3
155	TAPE G-H PIX 14	APC/E L=45	1	3
156	TAPE G-H PIX 14	TANK 3/4 L=60	0	3
157	TAPE G-H PIX 15	APC/E L=45	0	3
158	TAPE G-H PIX 15	TANK 3/4 L=60	0	4

PICT. NO.	COMMENTS	OR DESCRIPTION		POL	SHK
159	TAPE H-I	PIX 1 APC/E	L=40	0	3
160	TAPE H-I	PIX 1 TANK 3/4	L=70	0	5
161	TAPE H-I	PIX 2 2 1/2	L=90	0	7
162	TAPE H-I	PIX 3 APC/E	L=40	0	3
163	TAPE H-I	PIX 3 TANK 3/4	L=100	0	8
164	TAPE H-I	PIX 4 APC/E	L=40	0	3
165	TAPE H-I	PIX 4 TANK 3/4	L=60	0	5
166	TAPE H-I	PIX 5 APC/E	L=40	0	3
167	TAPE H-I	PIX 5 TANK 3/4	L=70	0	5
168	TAPE H-I	PIX 6 ?	L=10	0	1
169	TAPE H-I	PIX 6 TANK	L=40	0	3
170	TAPE H-I	PIX 7 ?	L=10	0	1
171	TAPE H-I	PIX 7 ?	L=8	0	1
172	TAPE H-I	PIX 7 TANK	L=20	0	2
173	TAPE H-I	PIX 8 APC/E	L=10	0	1
174	TAPE H-I	PIX 9 APC/E	L=8	0	1
175	TAPE H-I	PIX 10 APC/E	L=10	0	1
176	TAPE H-I	PIX 11 APC/E	L=10	0	1
177	TAPE H-I	PIX 12 APC/E	L=10	0	1
178	TAPE H-I	PIX 13 APC/E	L=5	0	1
179	TAPE H-I	PIX 13 TANK 3/4	L=5	0	1
180	TAPE H-I	PIX 14 TANK 3/4	L=70	0	6
181	TAPE H-I	PIX 15 APC/E	L=50	0	3

PICT. NO.	COMMENTS	OR DESCRIPTION		POL	SHK
204	TAPE L	PIX 1 APC/E	L=8	0	1
205	TAPE L	PIX 2 APC/E	L=5	0	1
206	TAPE L	PIX 2 TANK	L=10	0	1
207	TAPE L	PIX 3 APC/E	L=8	0	1
208	TAPE L	PIX 3 TANK	L=15	0	1
209	TAPE L	PIX 4 TANK	L=60	0	4
210	TAPE L	PIX 4 TRUCK	L=50	0	4
211	TAPE L	PIX 5 TANK 3/4	L=60	0	5
212	TAPE L	PIX 6 APC/E	L=40	0	2
213	TAPE L	PIX 7 APC/E	L=40	0	3
214	TAPE L	PIX 8 APC/E	L=40	0	2
215	TAPE L	PIX 9 TANK 3/4	L=60	0	4
216	TAPE L	PIX 10 APC/E	L=40	0	2

PICT. NO.	COMMENTS	OR DESCRIPTION		POL	SHK
217	TAPE H-N	PIX 1 APC	L=7	0	1
218	TAPE H-N	PIX 2 APC	L=10	0	1
219	TAPE H-N	PIX 3 APC	L=10	0	1
220	TAPE H-N	PIX 4 APC	L=10	0	1
221	TAPE H-N	PIX 5 APC	L=10	0	1
222	TAPE H-N	PIX 6 APC	L=7	0	1
223	TAPE H-N	PIX 7 TANK	L=25	0	2
224	TAPE H-N	PIX 7 APC	L=20	0	2
225	TAPE H-N	PIX 8 TANK	L=25	0	2
226	TAPE H-N	PIX 8 APC	L=20	0	2
227	TAPE H-N	PIX 9 TANK	L=30	0	2
228	TAPE H-N	PIX 10 TANK	L=30	0	2
229	TAPE H-N	PIX 10 APC	L=25	0	2
230	TAPE H-N	PIX 11 TANK	L=40	0	2
231	TAPE H-N	PIX 11 TANK RE-00	L=40	0	3
232	TAPE H-N	PIX 12 TANK	L=40	0	3
233	TAPE H-N	PIX 12 APC	L=20	0	2
234	TAPE H-N	PIX 13 TANK	L=40	0	2
235	TAPE H-N	PIX 14 TANK	L=40	0	3
236	TAPE H-N	PIX 14 APC	L=20	0	2
237	TAPE H-N	PIX 15 TANK	L=35	0	1
238	TAPE H-N	PIX 15 APC	L=15	0	1

PICT. NO.	COMMENTS	OR DESCRIPTION		POL	SHK
239	TAPE N-0	PIX 1 TRUCK	L=60	0	5
240	TAPE N-0	PIX 2 TANK/E	L=60	0	4
241	TAPE N-0	PIX 3 TRUCK	L=80	0	5
242	TAPE N-0	PIX 3 TANK	L=70	0	4
243	TAPE N-0	PIX 4 TRUCK	L=80	0	5
244	TAPE N-0	PIX 5 TANK/E	L=60	0	4
245	TAPE N-0	PIX 6 TANK/E	L=60	0	4
246	TAPE N-0	PIX 7 TRUCK	L=80	0	5
247	TAPE N-0	PIX 8 TRUCK	L=50	0	4
248	TAPE N-0	PIX 9 TRUCK	L=50	0	3
249	TAPE N-0	PIX 10 TRUCK	L=50	0	4
250	TAPE N-0	PIX 11 TRUCK	L=50	0	3
251	TAPE N-0	PIX 12 TRUCK	L=80	1	5
252	TAPE N-0	PIX 13 TANK	L=60	1	4
253	TAPE N-0	PIX 14 TRUCK	L=80	1	5

PICT. NO.	COMMENTS	OR DESCRIPTION		POL	SHK
254	TAPE J-K	PIX J-1 APC/E	L=50	0	3
255	TAPE J-K	PIX J-1 REPEAT		0	2
256	TAPE J-K	PIX J-2 TANK 3/4	L=70	0	4
257	TAPE J-K	PIX J-3 APC/E	L=50	0	3
258	TAPE J-K	PIX J-4 APC/E	L=50	0	3
259	TAPE J-K	PIX J-5 TANK 3/4	L=80	0	5
260	TAPE J-K	PIX J-6 APC/E	L=40	0	2
261	TAPE J-K	PIX J-7 APC/E	L=50	0	3
262	TAPE J-K	PIX J-8 APC/E	L=40	0	3
263	TAPE J-K	PIX J-9 TANK 3/4	L=70	0	4
264	TAPE J-K	PIX J-10 TRUCK/E	L=50	0	4
265	TAPE J-K	PIX K-1 APC/E	L=40	1	2
266	TAPE J-K	PIX K-2 APC/E	L=50	1	3
267	TAPE J-K	PIX K-3 APC	L=20	0	1
268	TAPE J-K	PIX K-4 TRUCK	L=50	0	4
269	TAPE J-K	PIX K-5 TANK 3/4	L=80	0	5

TABLE 3-II LIST OF DATA WINDOWS

PICT. NO.	TYPE	MULX	MULT	SEQUENCE	COMMENTS OR DESCRIPTION	POL	SHK
270		10	8	1	TAPE A PIX 7 FALSE ALARM WINDOWS	0	2
350		9	8	1	TAPE A PIX 10 FALSE ALARM WINDOWS	0	2
422		8	8	1	TAPE B PIX 2 FALSE ALARM WINDOWS	0	2
486		10	8	1	TAPE B PIX 9 F A EXCLUDE TGT	1	2
566		10	8	1	TAPE D PIX 9 F. A.	0	2
646		10	4	1	TAPE E PIX 10 F. A.	0	2
686		20	16	1	TAPE D PIX 9 F. A.	0	1

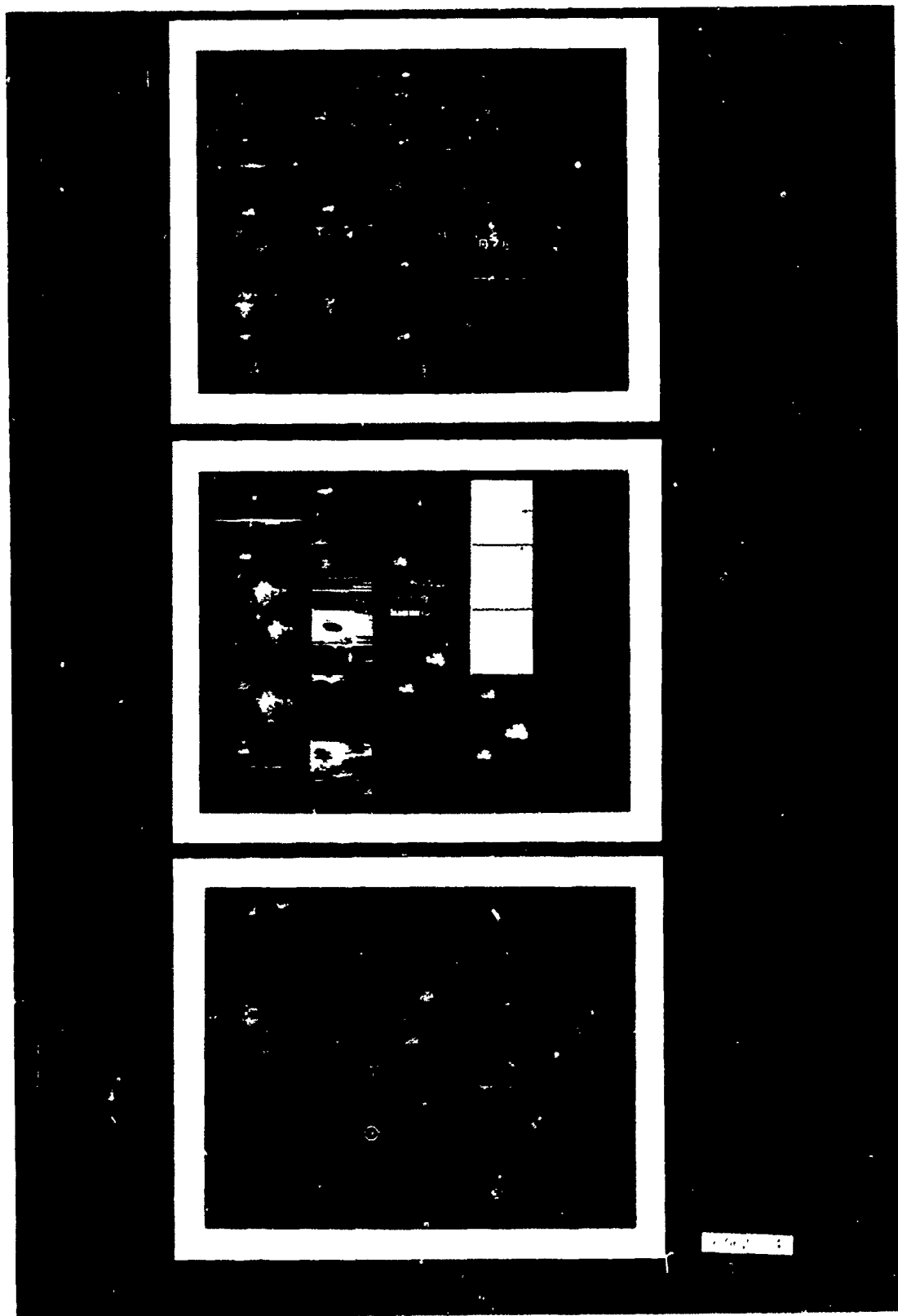


Figure 3-3. Photo Playbacks of 50 x 50 Images

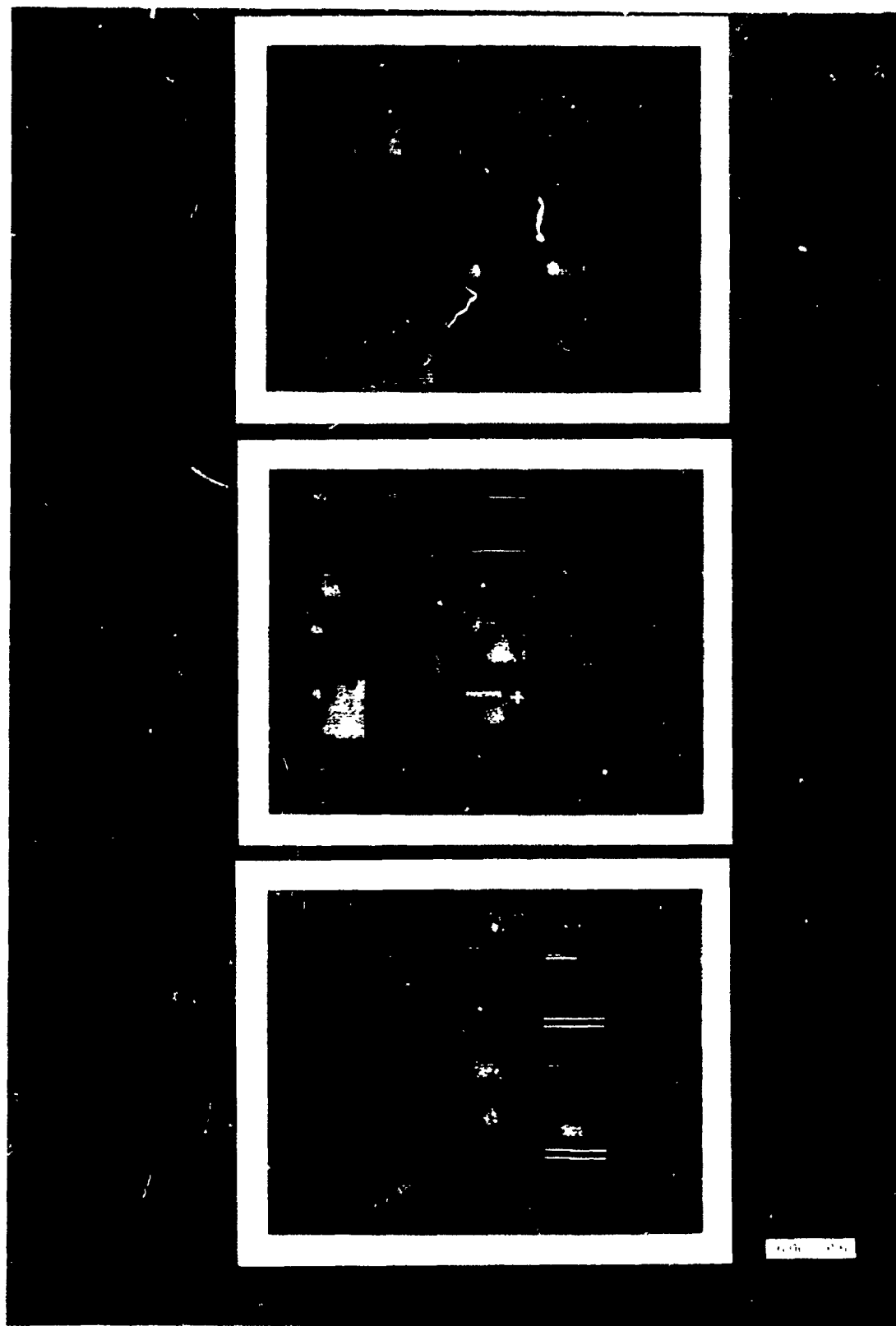


Figure 3-3. Photo Playbacks of 50 x 50 Images

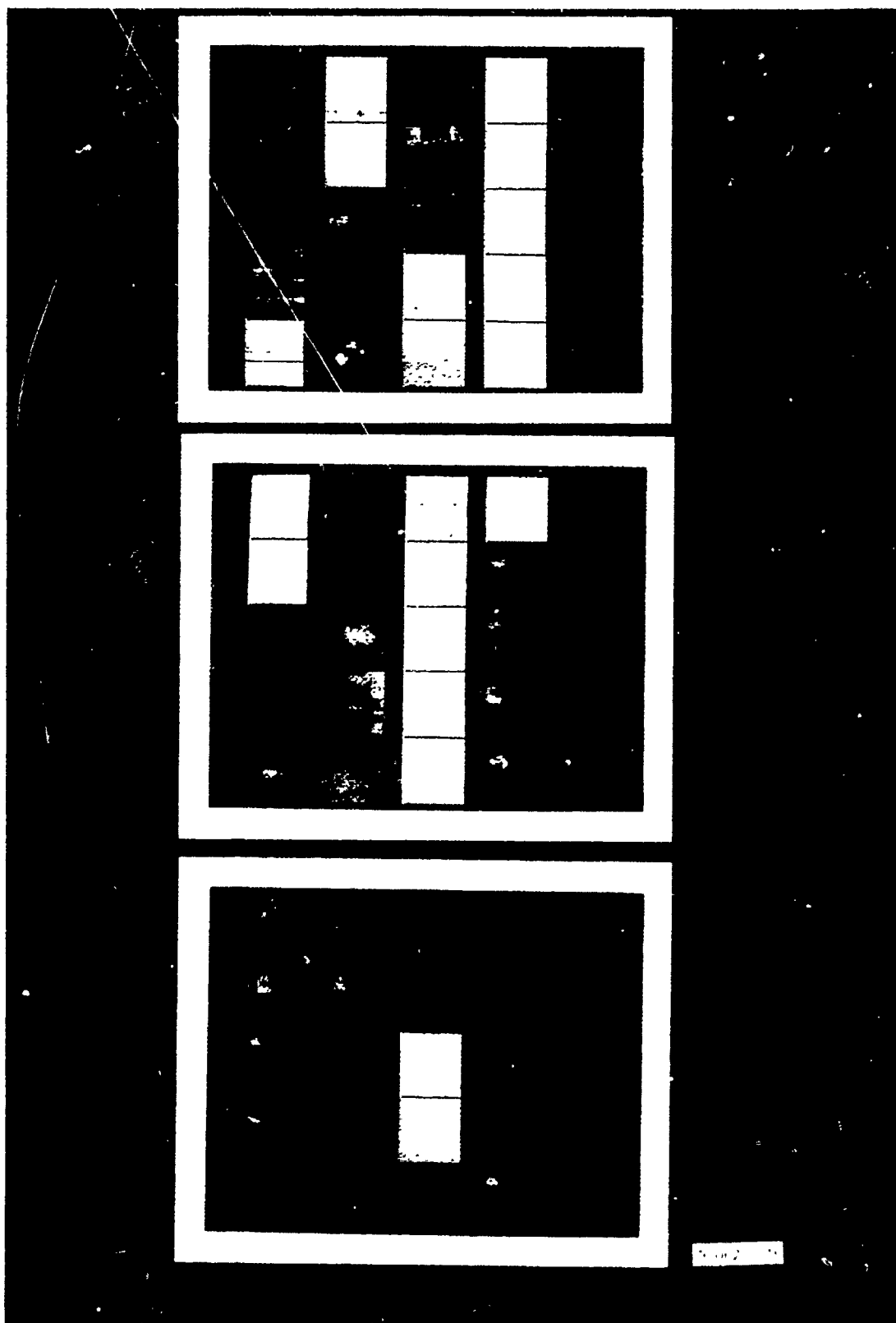


Figure 3-3. Photo Playbacks of 50 x 50 Images

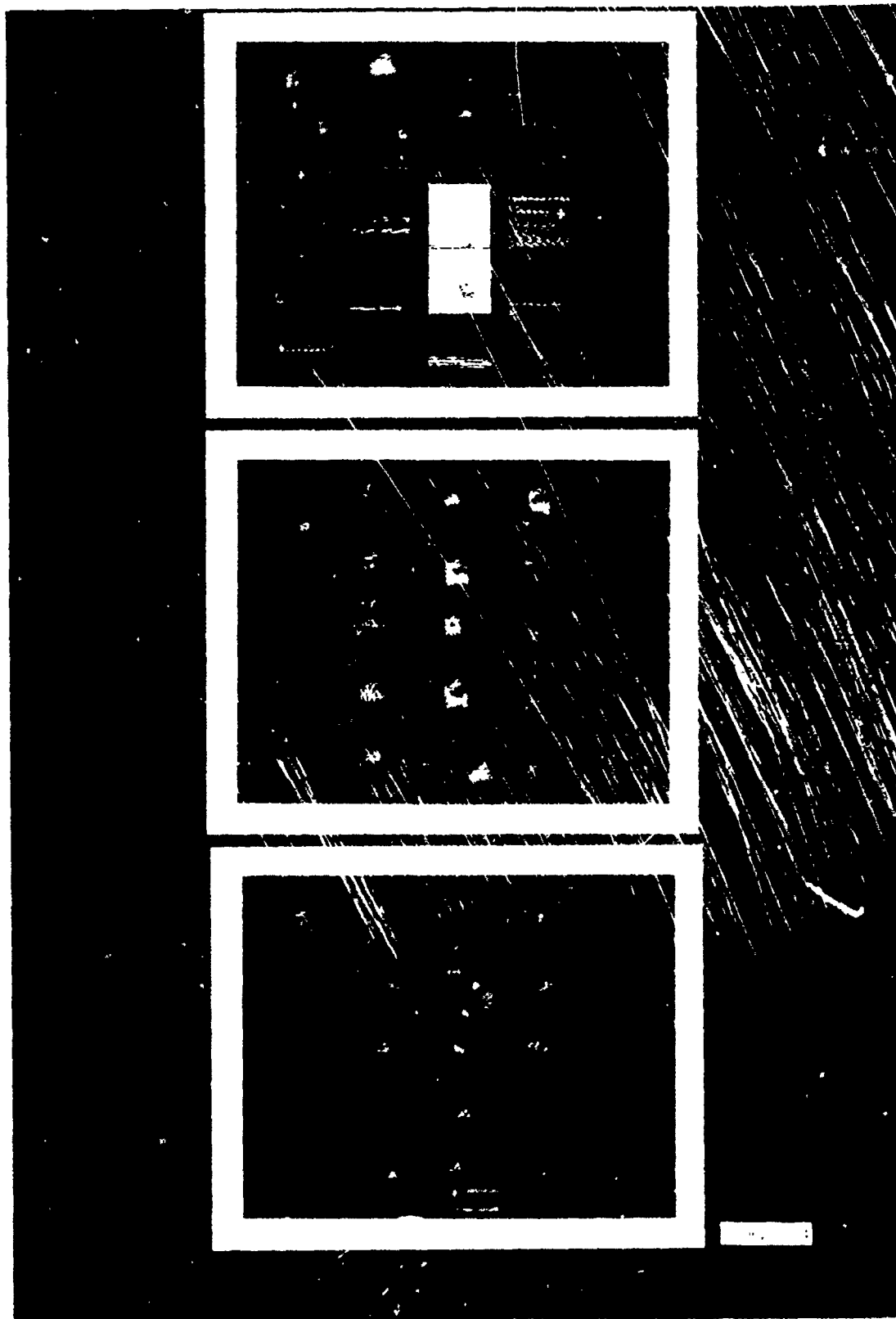


Figure 3-3. Photo Playbacks of 50 x 50 Images



Figure 3-3. Photo Playbacks of 50 x 50 Images

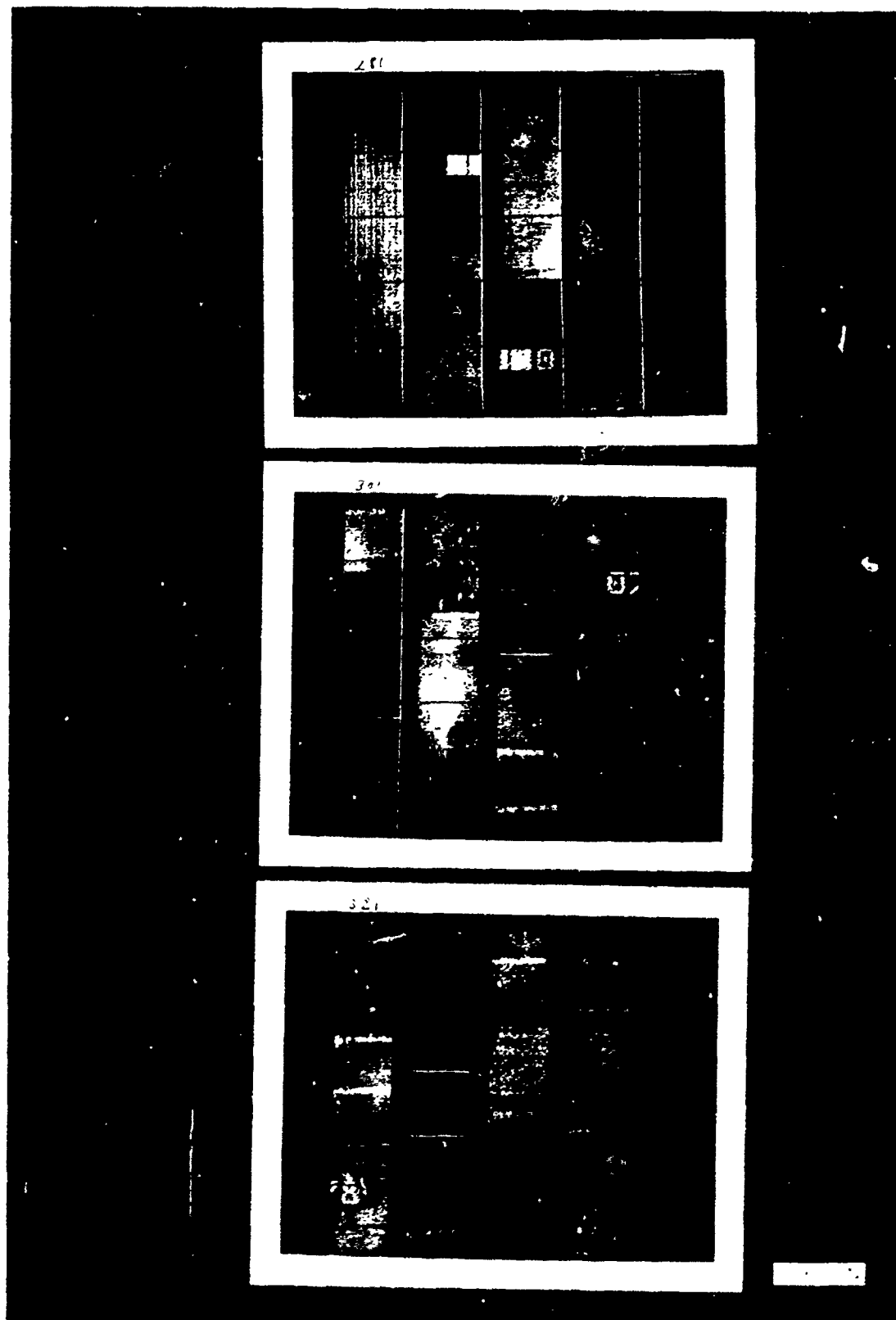


Figure 3-3. Photo Playbacks of 50 x 50 Images

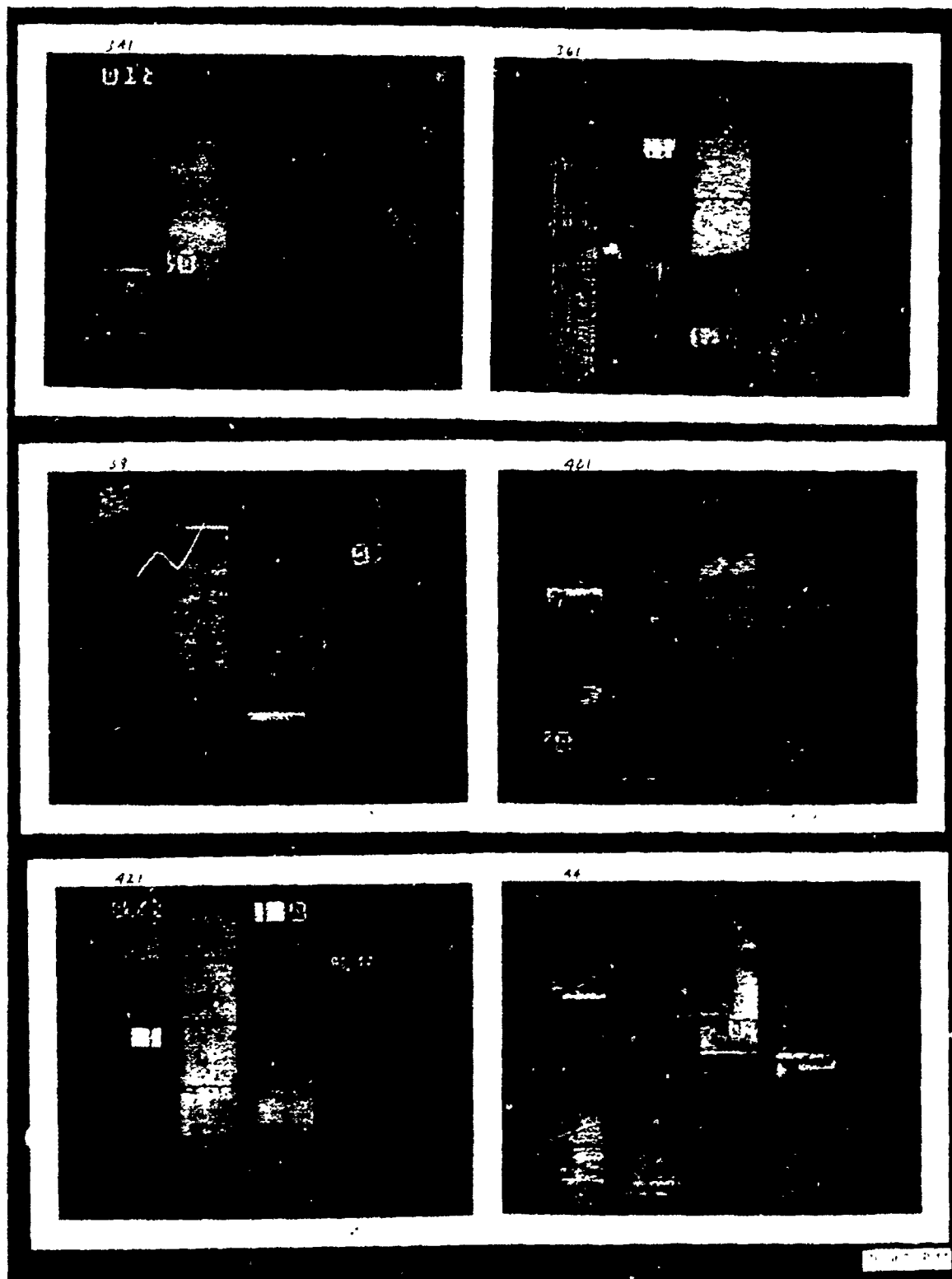


Figure 3-3. Photo Playbacks of 50 x 50 Images

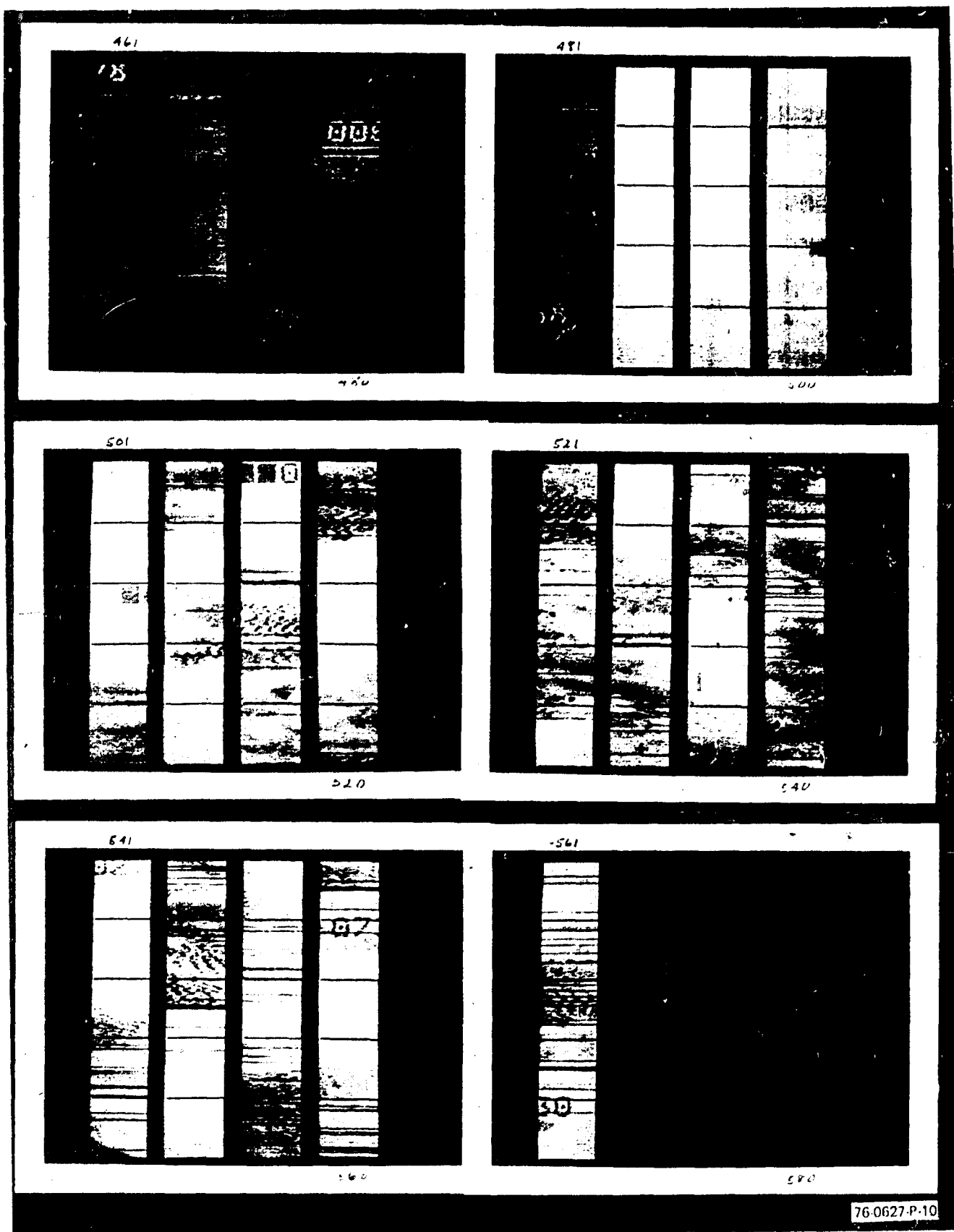


Figure 3-3. Photo Playbacks of 50 x 50 Images

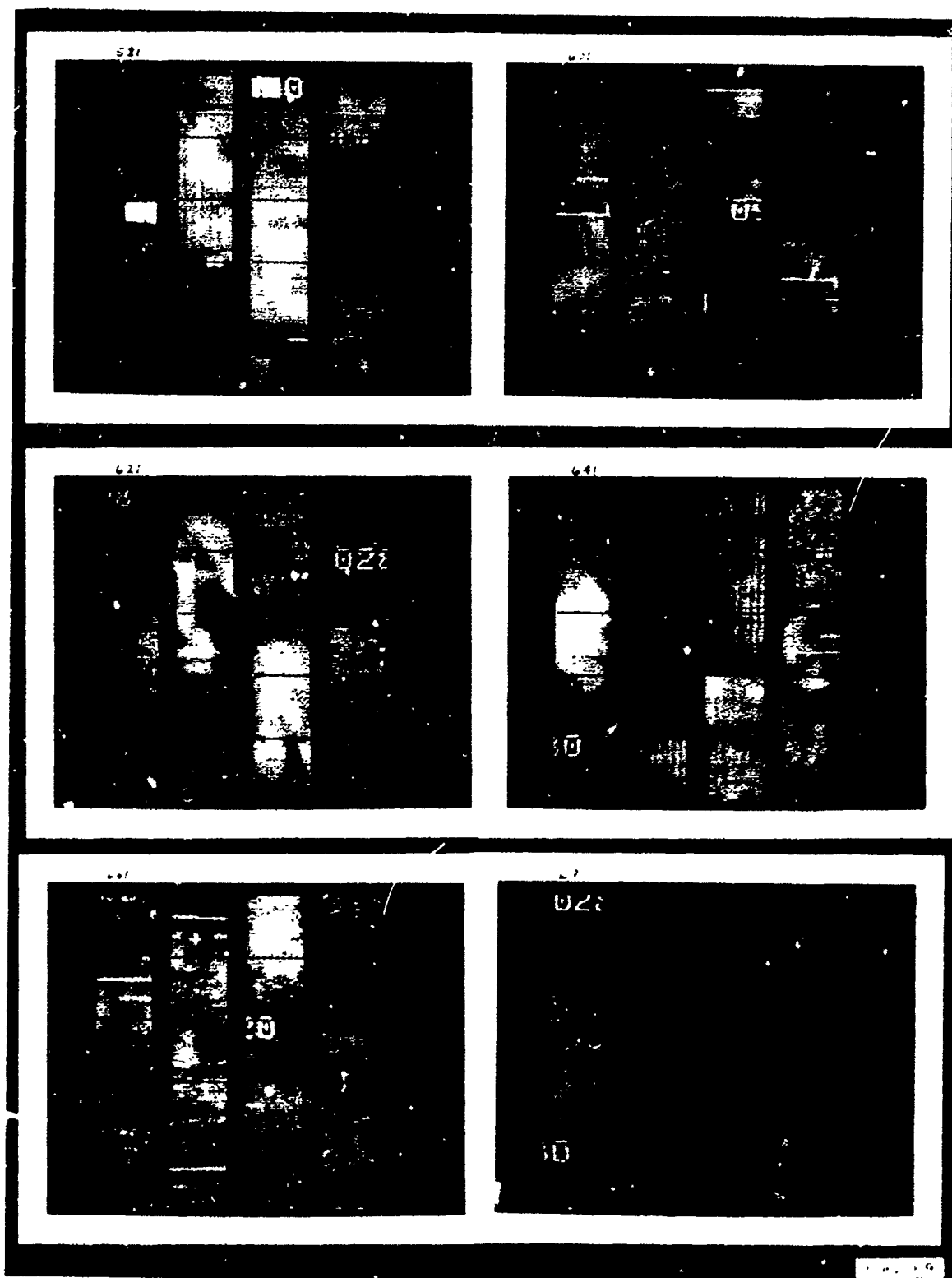


Figure 3-3. Photo Playbacks of 50 x 50 Images

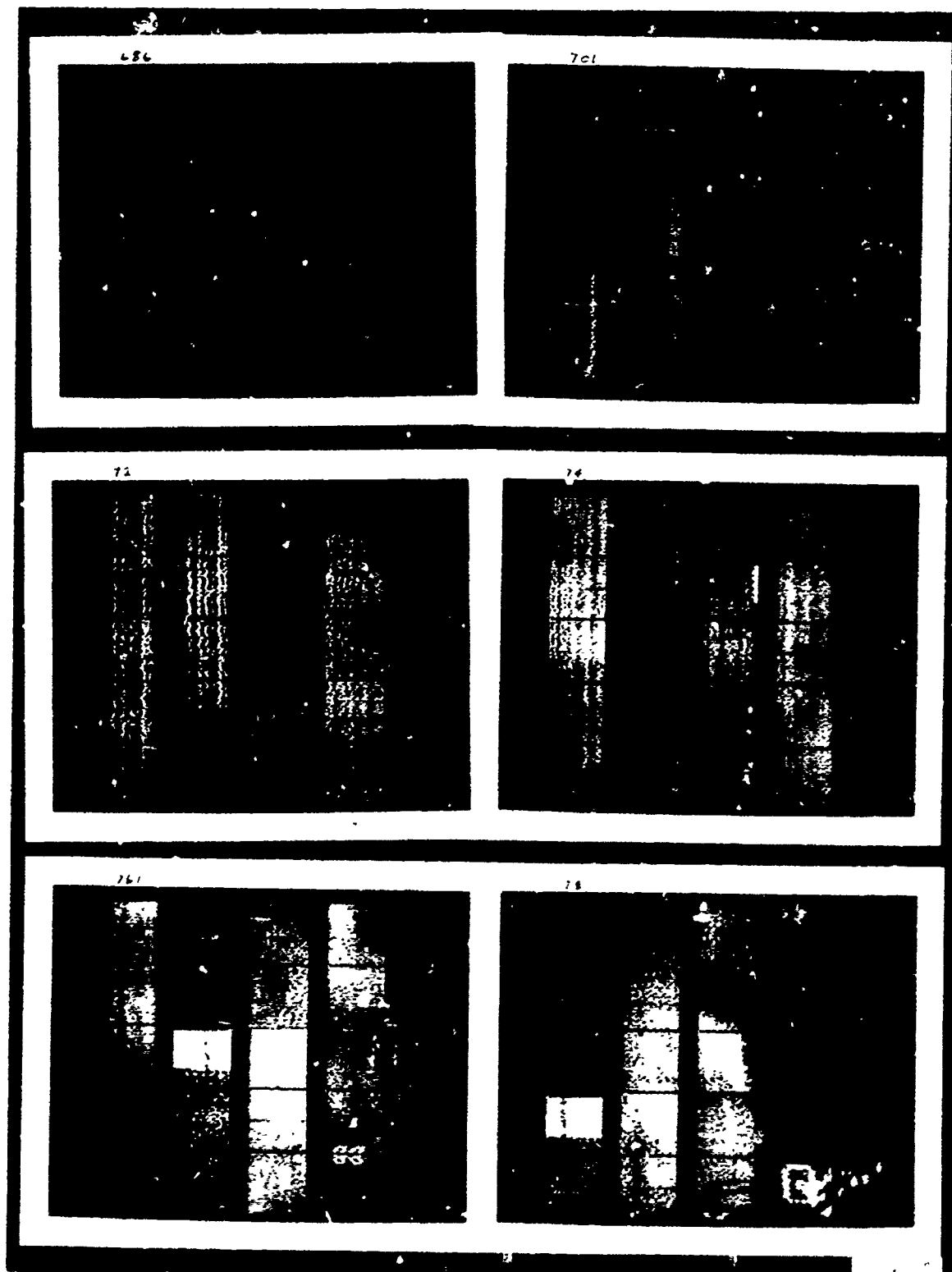


Figure 3-3. Photo Playbacks of 50 x 50 Images

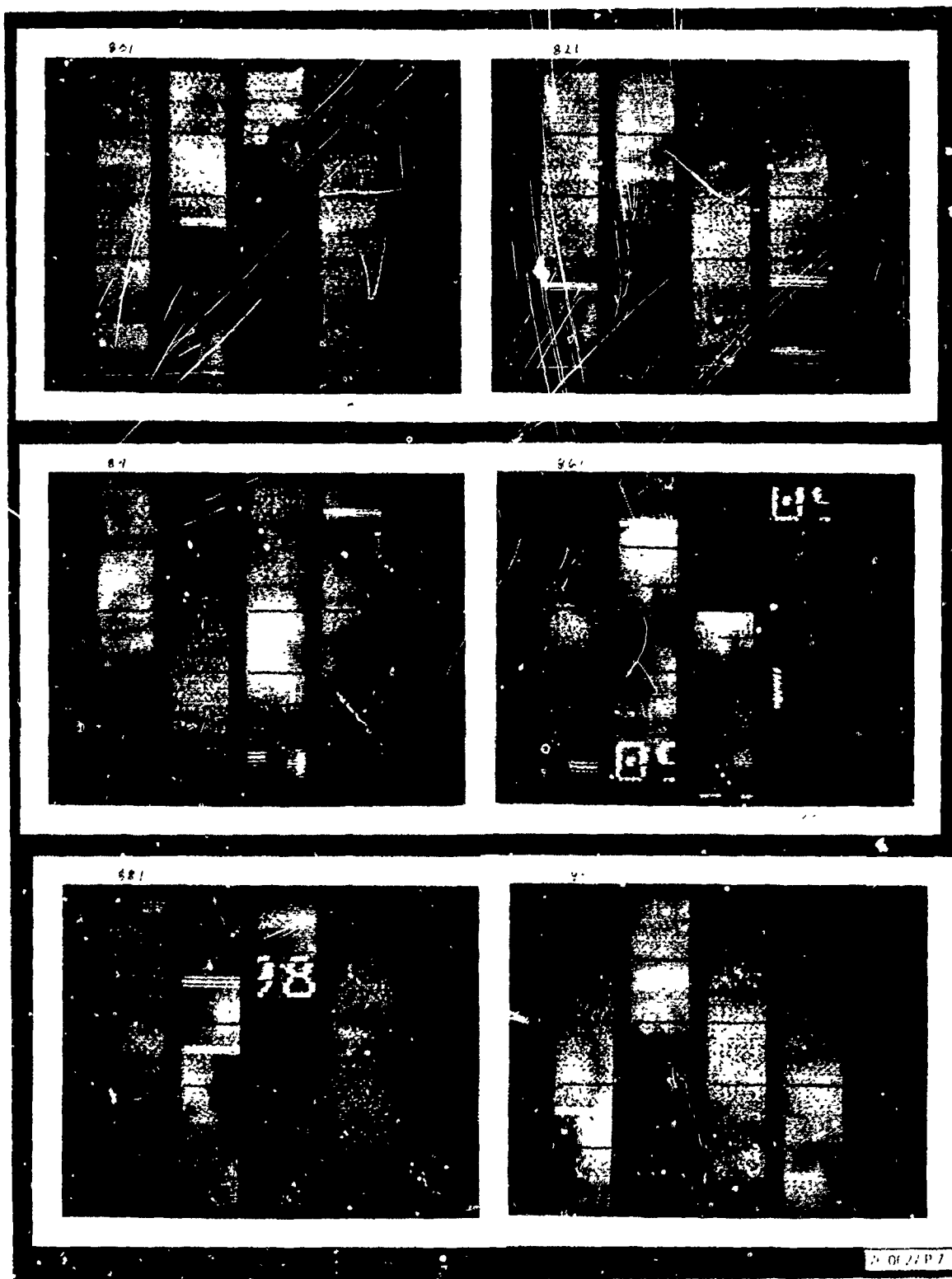


Figure 3-3. Photo Playbacks of 50 x 50 Images

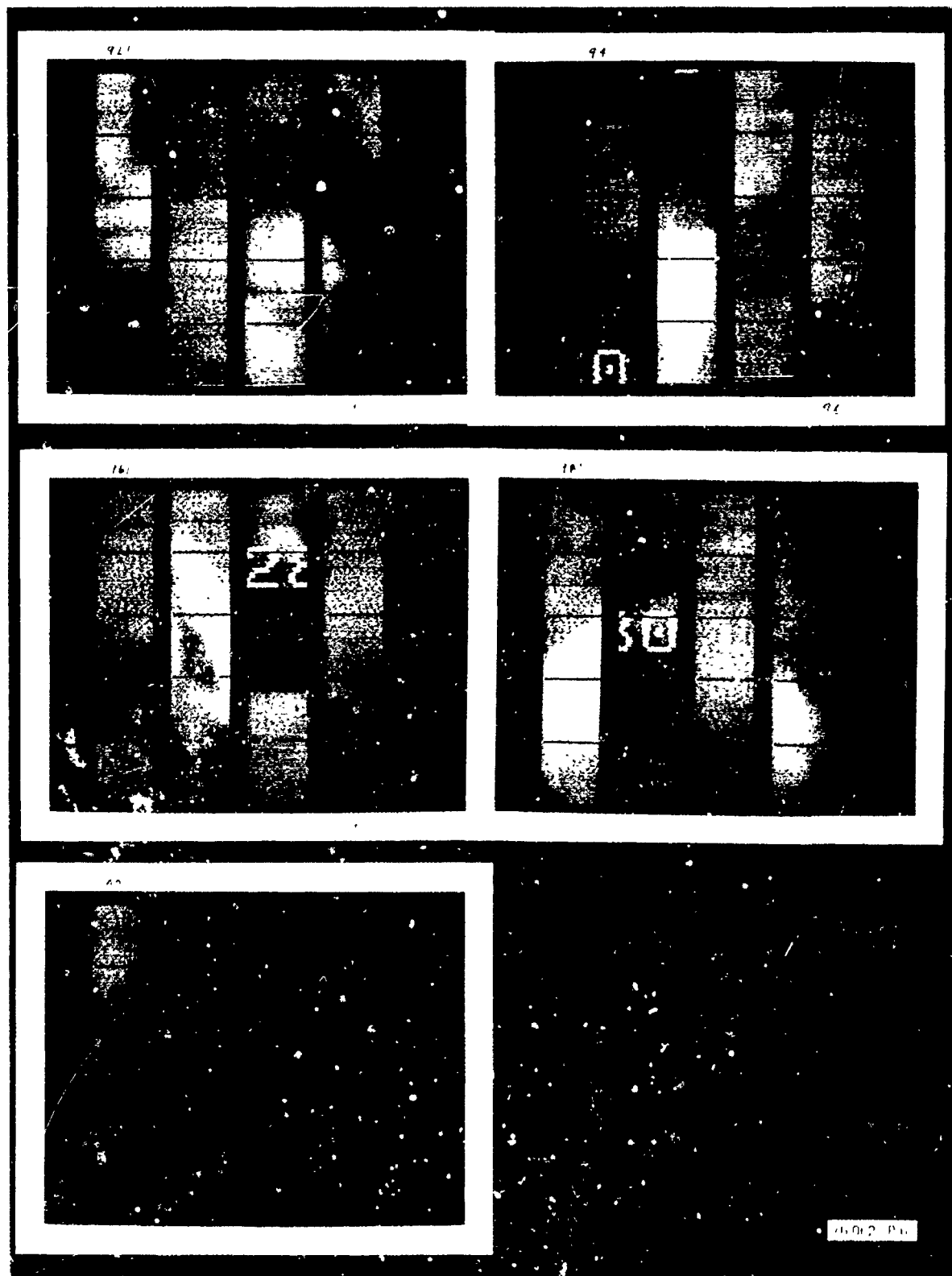


Figure 3-3. Photo Playbacks of 50 x 50 Images

Windows containing targets of unknown ground truth were excluded. Also, windows from Tape N-0 often did not contain a target, indicating incorrect correspondence with the film strips (as happened with Tape I-J). These samples were naturally excluded.

During the various examinations of the window playbacks, it was observed that there were many seriously degraded target images. A quick check showed that the problem was the ripple, herringbone, and Moiré distortion seen in the film strips and discussed earlier. It was especially apparent in the 50 x 50 window playbacks that were not shrunk because they are, in effect, a blown-up version of the original. Two examples of the degraded windows are shown in Figure 3-4. The upper 50 x 50 window contains a tank lifted from image L-2 (located by the arrow in the playback of L-2). The lower 50 x 50 window is an APC lifted from image L-3. Samples that had similar serious degradation were keyed for later reference.

Finally, Table 3-III shows the number of samples used in the training and test sets. The degraded samples were also used, but are tabulated separately in the table.\*

### 3.3 The Image Processing Sequence

Before proceeding into training, a brief review of the processing sequence is in order. Figure 3-5 shows the flow of data through the processor. Sub-images of size 50 x 50 pixels are lifted from the original scenes of size 800 x 1024. These small windows are written on another digital tape and photographically reconstructed.

---

\*The totals shown in Table 3-III differ slightly from those proposed in the December '75 Progress Report. The difference arises because later examination of the newly available playbacks indicated some missing targets, as already mentioned.

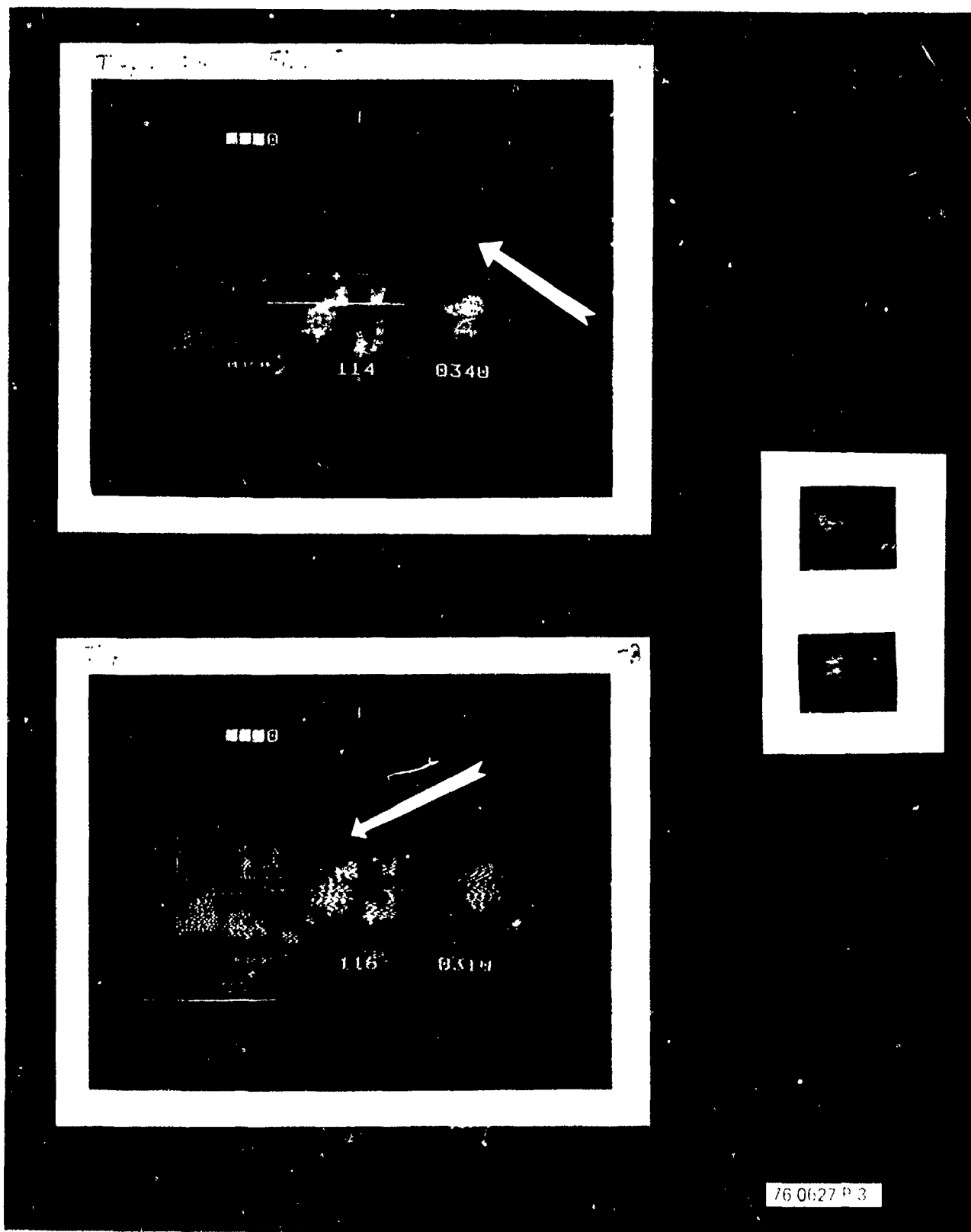


Figure 3-4. Two Examples of Degraded Samples

TABLE 3-III NUMBER OF SAMPLES

DATA BASE - NUMBER OF SAMPLES

	<u>TANK</u>	<u>TRUCK</u>	<u>APC</u>	<u>FALSE ALARM IMAGES</u>
TRAINING SET:				
ACCEPTABLE IMAGES	26	11	13	26
DEGRADED IMAGES	<u>15</u>	<u>8</u>	<u>18</u>	<u>      </u>
TOTAL	41	19	31	26
TEST SET:				
ACCEPTABLE IMAGES	30	9	11	213
DEGRADED IMAGES	<u>7</u>	<u>8</u>	<u>16</u>	<u>      </u>
TOTAL	37	17	27	213

The processing itself is split into two parts - a preprocessor "front-end", and a final processor. The preprocessor has an optional two-dimensional filter at the input. A single threshold is the only adjustable variable in the preprocessor. The level for this minimum gradient threshold was determined and set permanently prior to training. Final processing consists of two stages of false alarm screening, followed by a classification stage.

For estimating the performance of the system, scores were taken at the points indicated by the arrows. This will be further discussed in conjunction with the results tables.

### 3.4 Training Program

Since the digital image processing is split into two parts, it was convenient to perform the simulation and analysis of the training samples in two corresponding steps.

It was first necessary to select the amount of prefiltering and the level of minimum gradient in the preprocessor. A small set of windows from the training set were preprocessed using three different degrees of filtering and three levels of minimum gradient. Plots of the preprocessor outputs (subsets and blobs) were made on a Calcomp Model 763 plotter for visual analysis. It was evident that a 3 x 3 pixel 2-dimensional filter reduced the edge gradients on objects excessively. A 2 x 2 size filter was not excessive, yet it did provide some additional noise reduction.\*

The minimum gradient threshold determines the sensitivity of the preprocessor. As it is lowered, the number of subsets increases; i.e., fainter edges are allowed to come through. Thus, as the threshold is reduced, fainter

---

\*As described earlier, those windows that were shrunk were consequently being additionally filtered, as well as reduced in resolution.

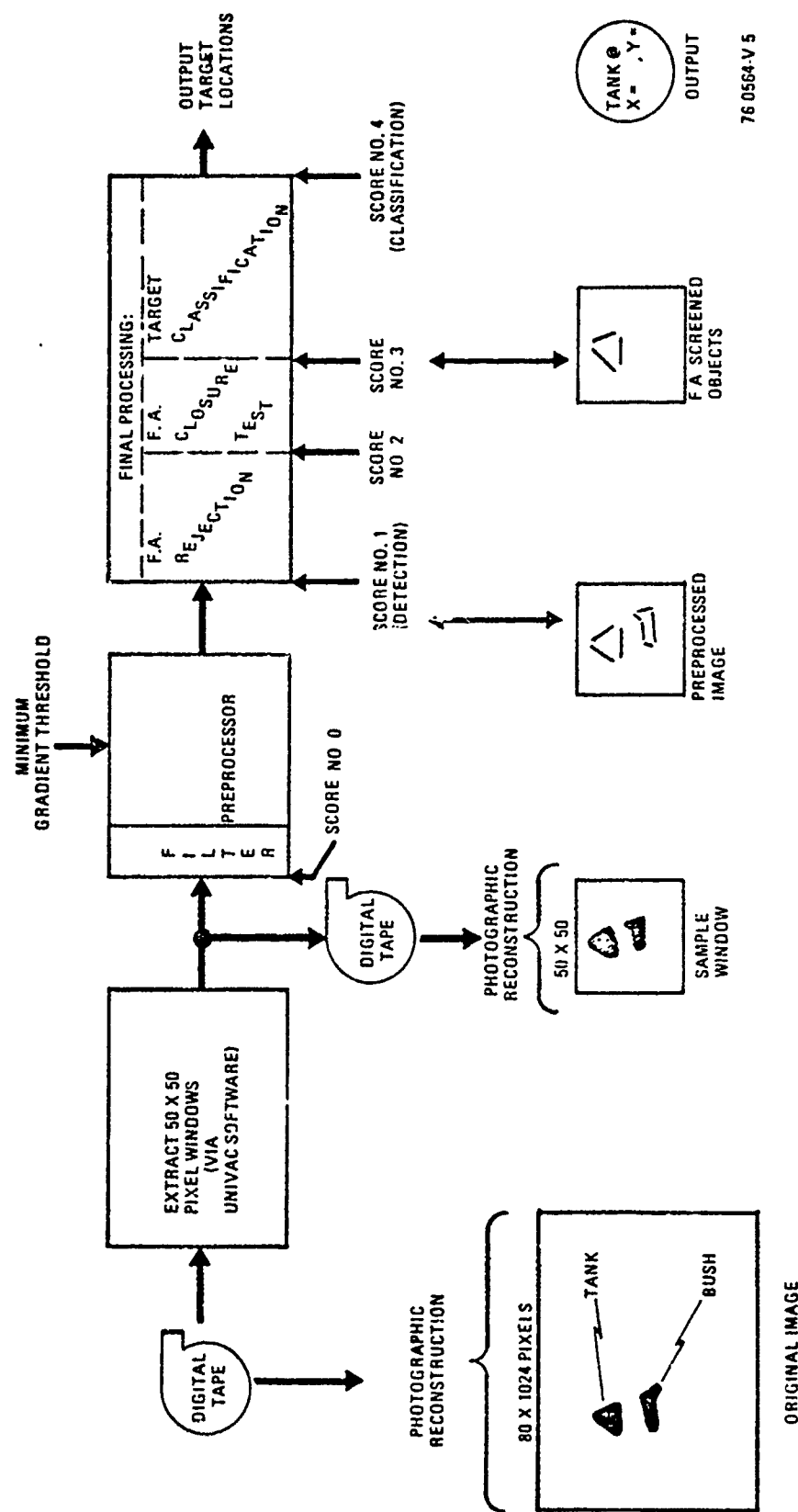


Figure 3-5 Data Flow Through Simulated Processor

targets will be detected. However, since the amount of clutter and noise data is also increasing, the ultimate false alarm rate will be higher.

To provide the greatest number of training patterns possible for feature analysis, including those of faint targets, a low minimum gradient threshold was desirable. The very lowest that had been run was 2.0\*. However, that setting provided too much background detail, which would interfere with the computation of the recognition features for training. Therefore, a value of 2.5 was chosen. Subsequent training and test runs were made at that threshold level.

The entire training set of windows were then preprocessed and the results saved on magnetic tape. From a previous program (Ref. 3) a set of recognition features had been developed for a 4 class environment (tank, jeep, truck, and personnel). Since this software already existed, an initial trial with these features was attempted. Specifically, the training set was processed through final classification using the existing program. Since only one half the target types were the same and one recognition feature was unavailable (range), the actual classification results were ignored. However, the values of the recognition features that were calculated and printed out were tabulated for each target sample. Scatterplots of these features were then made.

Experience from previous programs showed that the usual statistical measures such as means and variances can frequently be misleading because the distributions are often multimodal. Different target viewing angles, resolutions, etc. yield different modes. Analysis using scatterplots proved to be the most effective method to quickly determine separability of the classes.

---

\*That is, 2.0 out of 32 possible gray levels.

The scatterplots suggested that the existing features Aspect,  $\Delta R$ , NFP, and Closure should be retained. The features Khoriz, and Number of Quadrants Active should be modified. The remainder should be dropped and new features added.

Calcomp plots had been made of the preprocessor outputs for all the windows. Investigations of the Calcomp plots suggested some new trial features. These were programmed into the final processor simulation, and the training windows were rerun. Tables and scatterplots were then made of the new and modified features. An analysis of the results indicated that some further modifications of the new features were needed.

After re-programming the modifications the training set was again rerun through the final processor simulation. Scatterplots of the new features were made. The plots suggested that the classes were not linearly separable. Therefore any training algorithm that did not converge unless there was separability should not be used.

As in previous studies, the number of samples for training is much too small to try parametric methods of designing a classifier, even if a distribution could be assumed. Nonparametric classifiers that require storage and searches of templates or sample patterns (e.g., k-nearest neighbor algorithms) are either too time consuming for real-time data rates or too limited in the number of models to handle all the variations in aspect angle, etc.

An adaptive training algorithm such as the sum-line algorithm that had been programmed in-house would be appropriate for a classifier. However, time did not permit experimentation with it under a multi-class condition. To expedite estimates of system performance, a two-layer classifier (Ref. 4)

was chosen. Subdecisions are made on the basis of deterministically designed boundaries from the scatterplots and the special features KHOLE and LONTOP. The subdecisions are then used in making a final decision.

Consequently, training was done by drawing decision boundaries from the scatterplots that would minimize the final error-rate and still provide reasonable extrapolative performance. The scatterplots of the training set data (target windows) are shown in Figures 3-6(a), (b), and (c). The decision boundaries have also been drawn in. Note that the region NON-TRUCK actually means APC and TANK. Linear boundaries are used because they are simple to implement in software (e.g., in a  $\mu$ -Processor) and are computationally fast.

The final boundaries for the various features are shown in Figure 2-11 of Section 2.2. The final decision of the target class is made by taking a majority vote of the outcome of the individual boundary sets. In case of a tie, additional rules were developed from investigations of the scatterplots. The final decision logic is also described in Section 2. A tabulation of the features, grouped by target class, suggested several additional criteria to separate the target features from non-target (or false alarm) features. These tests are shown in Table 2-I.

### 3.5 Scoring of Training Set

Upon completion of the classification algorithm design, the training samples were processed and scored. Table 3-IV gives the results. The scores for the previously keyed degraded samples are separated from those of the remaining samples. An "\*" denotes the degraded sample scores. The leftmost column shows the number of target samples that were processed in this training set. Referring back to Figure 3-5, the number of targets at the output of

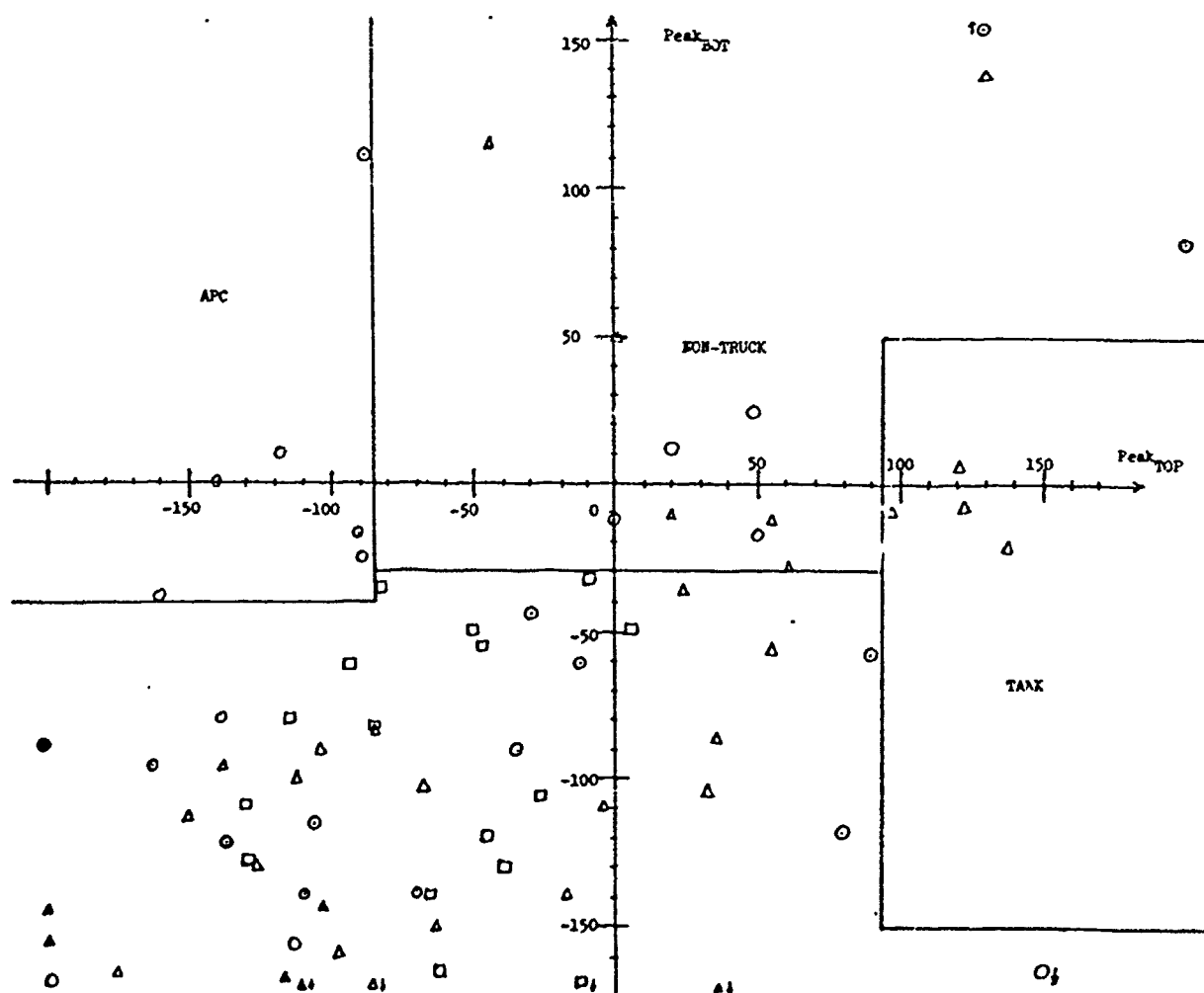
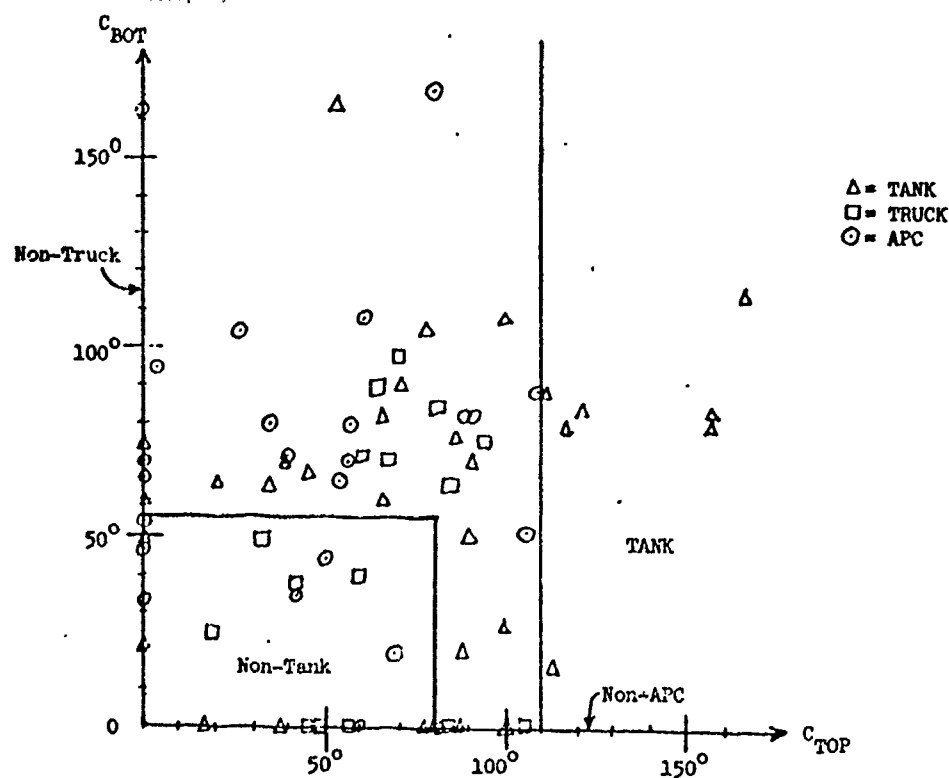


Figure 3-6(a). Training Set Scatterplots.

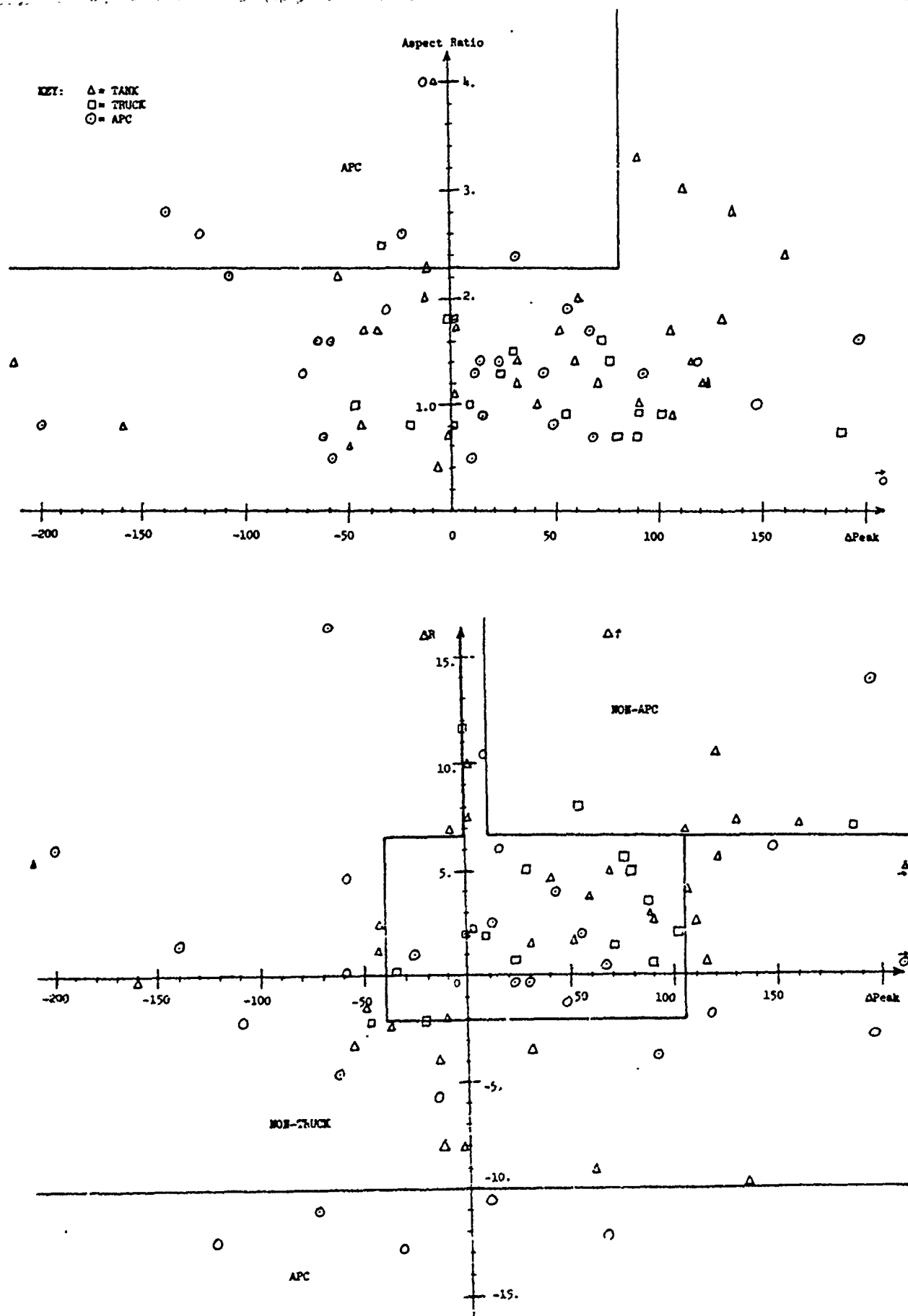


Figure 3-6(b). Training Set Scatterplots.

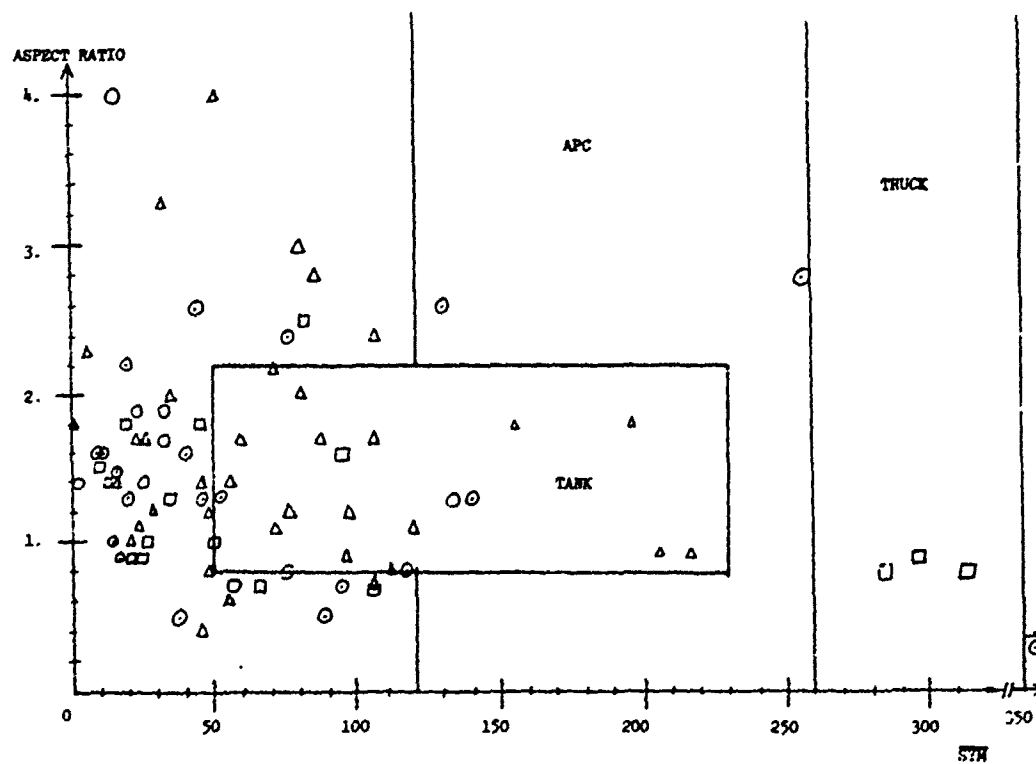
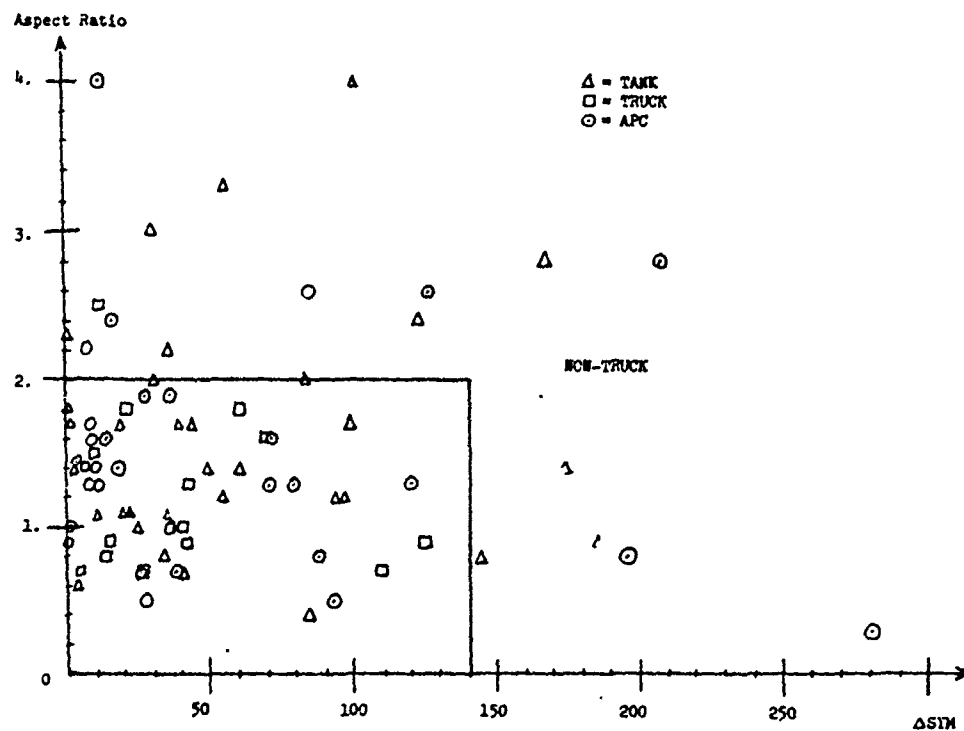


Figure 3-6(c). Training Set Scatterplots.

TABLE 3-IV TRAINING SET RESULTS -- RAW SCORES

	<u>NUMBER OF SAMPLES</u>	<u>NUMBER DETECTED</u>	<u>NUMBER REMAINING THRU REJECTION</u>	<u>NUMBER REMAINING THRU CLOSURE</u>	<u>NUMBER CORRECTLY CLASSIFIED</u>	<u>NUMBER FROM COLUMN 2 CORRECTLY CLASSIFIED</u>
TANKS	26 + 15*	26 + 12*	25 + 10*	24 + 7*	21 + 4*	23 + 8*
TRUCKS	11 + 8*	11 + 5*	10 + 5*	9 + 1*	6 + 0*	8 + 4*
APC	<u>13 + 18*</u>	<u>13 + 13*</u>	<u>12 + 8*</u>	<u>11 + 2*</u>	<u>10 + 2*</u>	<u>10 + 10*</u>
TOTAL	50 + 41*	50 + 30*	47 + 23*	44 + 10*	37 + 6*	41 + 22*

\*Denotes Degraded Images

the preprocessor ("Score #1" arrow) are tabulated in the second column of Table 3-IV. It is at this point that the target has been initially detected.

The final processor has two stages of false alarm screening. The number of targets remaining after the first stage (see Figure 3-5, "Score #2" arrow) are listed in the third column of Table 3-IV. Those targets remaining after the "false alarm closure test" ("Score #3" arrow) are counted and given in the "Number Remaining Thru Closure" column. This completes the false alarm screening. All remaining objects are now assigned to a target class by the classification logic. Those targets that are correctly classified are counted ("Score #4") and listed in column 5 of the Table. A special count was also taken at the "Score #4" location. The computer simulation actually provided classification of all detected targets (i.e., bypassing false alarm screening). Scoring the classification of all detected targets will give a better estimate of how well the classification algorithm, itself, is performing, regardless of the screening performance. The righthand column of Table 3-IV gives this count.

A look at the data in Table 3-IV shows that the detection count is high, decreasing somewhat through the false alarm screening stages. It is apparent that the degraded samples do not perform nearly as well as the acceptable samples. While the closure test and classification stage take a heavy toll on the degraded samples, the acceptable samples do very well. Recalling the tremendous distortions, etc. of the degraded images, it is not at all surprising that they do not perform well.

The training results are shown in terms of performance percentages in Table 3-V. The "Combined Samples" column contains the score for the degraded and the acceptable samples combined together. Definitions for the different percentages is given below.

TABLE 3-V TRAINING SET RESULTS - PERCENTAGES

	<u>COMBINED SAMPLES</u>	<u>DEGRADED SAMPLES</u>	<u>ACCEPTABLE SAMPLES</u>
DETECTION	$\frac{80}{91} = 88\%$	$\frac{30}{41} = 73\%$	$\frac{50}{50} = 100\%$
DETECTED AND SCREENED	$\frac{54}{91} = 59\%$	$\frac{10}{41} = 24\%$	$\frac{44}{50} = 88\%$
CLASSIFICATION OF DETECTED AND SCREENED SAMPLES	$\frac{43}{54} = 80\%$	$\frac{6}{10} = 60\%$	$\frac{37}{44} = 84\%$
DETECTED, SCREENED, AND CLASSIFIED	$\frac{43}{91} = 47\%$	$\frac{6}{41} = 15\%$	$\frac{37}{50} = 74\%$
CLASSIFICATION PERFORMANCE	$\frac{63}{80} = 79\%$	$\frac{22}{30} = 73\%$	$\frac{41}{50} = 82\%$

DETECTION: This is the percentage of ALL target samples that were detected by the preprocessor, i.e., Score #1/Score #0 of Figure 3-5.

DETECTED AND SCREENED: This is the percentage of ALL target samples that remained through false alarm screening, i.e., Score #3/Score #0 of Figure 3-5. These remaining samples will all be next assigned one of the target classes. From an operational aspect and a human factors aspect, it is this score that is often termed "detection".

CLASSIFICATION OF DETECTED AND SCREENED SAMPLES: This is the percentage of the above target samples that were correctly classified, i.e., Score #4/Score #3 of Figure 3-5.

DETECTED, SCREENED, AND CLASSIFIED: This is the percentage of ALL target samples that were detected, screened, and correctly classified, i.e., Score #4/Score #0.

CLASSIFIER PERFORMANCE: As described earlier, a special count at location "Score #4" was taken to provide a performance estimate of the classification algorithms, independent of the false alarm screening. Specifically, all targets at location "Score #1" were run through the classification logic. We thus have the performance estimate:

Special Score #4/Score #1.

Referring back to the data in Table 3-V, it is verified that the detection rate is quite high, especially for the acceptable sample category. The Detected and Screened score is also good for the acceptable samples.

The degraded samples are too broken up by distortion to effectively pass through the screening stage, and therefore pull down the average score. Classification of the Detected and Screened targets is good. Even the degraded samples have a Passable score, nearly twice as good as random chance (33% for three equal target classes).

The Detected, Screened, and Classified rate is the product of the two scores above it. So naturally, it is lower than either score. It is evident that the lower performance of the degraded samples pulls down the "combined samples" score. Otherwise, the acceptable samples perform well. Finally, the Classifier Performance percentages show creditable performance, even on the degraded samples.

In addition to scoring the performance on target samples, several non-target images were included in the training set, for false alarm estimates. These windows were specifically chosen to include many target-like objects, more than an average scene would contain. This helped to derive more effective false alarm screening criteria in the training process.

The non-target windows were processed along with the target samples and scored. Initially, 14 out of 26 windows had false alarms. However, a detailed examination of the results showed that 6 of the false alarms were from extraneous image data, not a part of the FLIR scene. One source was the alphanumeric characters superimposed on the video. Other extraneous sources were black, and white horizontal lines through the image. These are not part of the FLIR video, but are from digitizing or magnetic tape errors. Figure 3-7 shows an example of a white line through the image. The upper picture is a playback of image J-9 on Tape I-J. The lower picture is a playback of J-9 on Tape

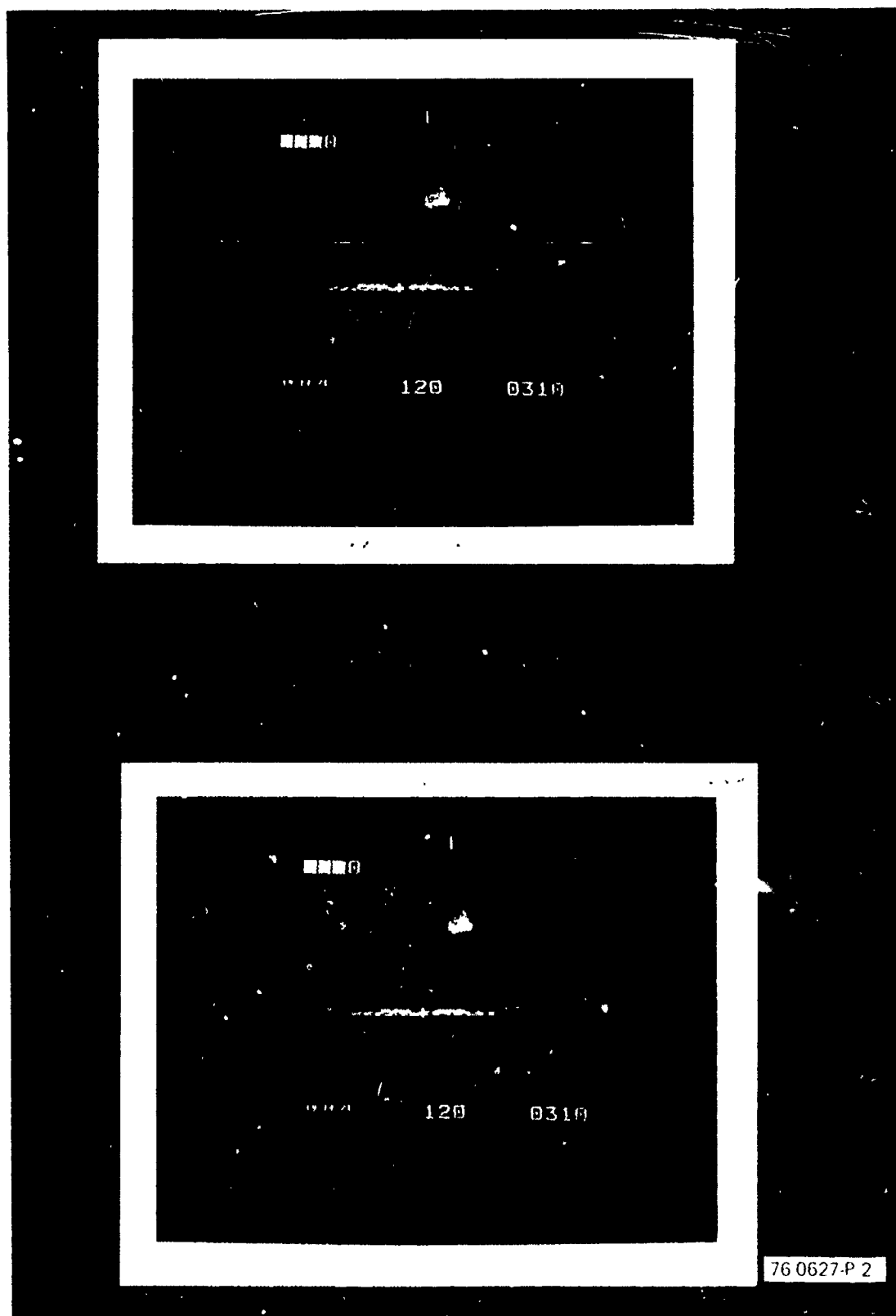


Figure 3-7. Extraneous White Line in Image

J-K; it does not contain the white line even though it's the same scene. The last extraneous causes of false alarms were the cursor and horizon line. These also would not normally be a part of the video sent to a processor.

Therefore, false alarms caused by these sources were henceforth excluded. That leaves 8 out of 26 windows with false alarms, or 31%. As indicated previously, an average scene would not contain as many targetlike objects over the whole field-of-view. Thus the average rate would be lower.

In addition, a higher preprocessor minimum gradient setting would reduce false alarms. A sensitive setting of 2.5 had been used to provide a greater number of target patterns for training purposes. For comparison, 61 non-target windows (including the previous 26) were processed at a preprocessor gradient level of 4.0. The false alarm rate then jumped down to  $8/61 = 13\%$ . Time did not permit re-processing the target windows to estimate the natural drop in detection or recognition rates. However, experience on a similar previous study indicated that the false alarm rate dropped much faster than the detection, or recognition rates for an increase in gradient setting.

A second control over the false alarm rate is in the final processor - the false alarm closure test. A variation of this parameter will be described in the next section.

### 3.6 Test Set Performance

Following the conclusion of the training phase, the test samples were processed using the thresholds and algorithms established during training. The same types of scores were then counted. Table 3-VI presents the raw data for the test set. As before, the degraded images have been separated

TABLE 3- VI TEST SET RESULTS - RAW SCORES

	<u>NUMBER OF SAMPLES</u>	<u>NUMBER DETECTED</u>	<u>NUMBER REMAINING THRU REJECTION</u>	<u>NUMBER REMAINING THRU CLOSURE</u>	<u>NUMBER CORRECTLY CLASSIFIED</u>	<u>NUMBER FROM COLUMN 2 CORRECTLY CLASSIFIED</u>
TANKS	30 + 7*	28 + 7*	24 + 5*	23 + 3*	19 + 3*	24 + 5*
TRUCKS	9 + 8*	7 + 8*	5 + 6*	4 + 3*	2 + 1*	2 + 1*
APC	<u>11 + 16*</u>	<u>11 + 16*</u>	<u>9 + 14*</u>	<u>9 + 4*</u>	<u>3 + 3*</u>	<u>7 + 6*</u>
TOTAL	50 + 31*	46 + 31*	38 + 25*	36 + 10*	24 + 7*	33 + 12*

\*Denotes Degraded Images

and marked by an "\*". The columns are the same as used in Table 3-IV. The results are more easily viewed as percentages, given in Table 3-VII. The definitions of the various scores are also the same as detailed earlier.

The Detection rate shown in the table is again excellent. The detected and Screened rate is good for the acceptable samples, but is inferior for the degraded samples. Similar results were experienced with the training set. Classification of Detected and Screened samples was lower than the training set, although still twice as high as random chance. This would indicate that additional samples should be used in the training set to derive more general classification boundaries. A look back at the raw data shows that the truck class was the main cause of the lower score. An examination of the scatter-plots for the truck class point out that the number of truck samples in both training and test is small. Therefore the full spread of their probable feature distributions was not well represented. A larger training set should provide better results.

The "product" score - Detected, Screened, and Classified was driven lower than the training scores by the lower classification performance. The Classification Performance was lower than the training set. The difference arises from the same problem as encountered by the Classification of Detected and Screened Samples score and discussed above. Additional training samples, especially for the truck class, should improve the performance.

A summary of the training and test results, by window number, is shown in Table VIII. The windows that were considered degraded in quality are indicated in the fourth column. In the "Result" column, a "C" indicates that the target was detected, screened, and correctly classified. If the target

TABLE 3-VII TEST SET RESULTS - PERCENTAGES

	COMBINED SAMPLES	DEGRADED SAMPLES	ACCEPTABLE SAMPLES
DETECTION	$\frac{77}{81} = 95\%$	$\frac{31}{31} = 100\%$	$\frac{46}{50} = 92\%$
DETECTED AND SCREENED	$\frac{46}{81} = 57\%$	$\frac{10}{31} = 32\%$	$\frac{36}{50} = 72\%$
CLASSIFICATION OF DETECTED AND SCREENED SAMPLES	$\frac{31}{46} = 67\%$	$\frac{7}{10} = 70\%$	$\frac{24}{36} = 67\%$
DETECTED, SCREENED, AND CLASSIFIED	$\frac{31}{81} = 38\%$	$\frac{7}{31} = 23\%$	$\frac{24}{50} = 48\%$
CLASSIFICATION PERFORMANCE	$\frac{45}{77} = 58\%$	$\frac{12}{31} = 39\%$	$\frac{33}{46} = 72\%$

TABLE 3-VIII. TRAINING AND TEST RESULTS - BY WINDOW

<u>Window Number</u>	<u>Training</u>	<u>Test</u>	<u>Degraded?</u>	<u>Result</u>
1	X		X	C
2	X			C
3	X			C
4		X	X	D
5		X		M
6	X		X	D
8		X		C
9	X			C
11	X			C
12		X		C
13		X	X	C
14	X			C
15	X			C
17		X		M
20		X		M
21	X			C
22		X		D
25	X			C
26	X			C
27	X			M
31		X		C
32		X		C
33	X			C
34	X			D
35	X			D
36	X			C
37	X			C
38		X		C
39		X		D
40		X		D
41	X			C
43	X		X	M
44	X			C
45	X			C
46		X		M
47		X		C
48		X		C
49	X		X	M
51		X		M
52		X		M
53		X		C
54	X		X	M
55	X		X	M
57		X	X	M
58		X		M

KEY: C = Detected, Screened, and Correctly Classified  
D = Detected and Screened, but Misclassified  
M = Missed (No Detection)

TABLE 3-VIII. TRAINING AND TEST RESULTS - BY WINDOW (Continued)

<u>Window Number</u>	<u>Training</u>	<u>Test</u>	<u>Degraded?</u>	<u>Result</u>
60	X		X	M
61	X			C
62		X	X	M
63		X		M
64		X		C
65	X		X	M
67		X		M
68	X			M
69	X		X	C
73	X		X	M
74		X		D
75	X			C
76		X		C
77	X			C
79		X		D
80	X		X	M
81	X			C
82	X			C
83		X		M
84	X			C
85		X		D
86		X		C
91	X		X	D
92	X		X	M
95		X		D
96		X		D
97		X	X	M
98	X		X	M
99	X		X	C
100	X			M
101	X		X	C
102	X			M
103	X		X	M
104		X		M
105		X		M
106		X	X	M
107	X		X	M
108	X		X	D
109	X		X	M
111		X	X	M
112		X	X	C
113		X	X	C
114		X	X	M
115		X	X	M
116		X	X	C
117	X		X	M
118	X		X	M
119	X			C
120		X	X	D
121		X	X	C
122		X		M
126	X			C
127	X			C
128	X		X	C
130		X	X	M

TABLE 3-VIII. TRAINING AND TEST RESULTS - BY WINDOW (Continued)

<u>Window Number</u>	<u>Training</u>	<u>Test</u>	<u>Degraded?</u>	<u>Result</u>
132	X		X	C
136		X		M
139	X		X	M
145	X		X	M
148		X	X	M
151	X			C
152		X		C
153	X			C
154	X			D
155		X		C
156		X		C
157		X		C
158		X		C
159	X			C
160	X			C
161	X			D
162		X	X	M
163		X		C
164	X			D
165	X			C
166		X		D
167		X		C
169		X	X	C
172	X		X	M
174	X		X	M
175		X	X	D
180		X		C
181		X		D
204		X	X	C
205	X		X	M
206	X		X	M
207		X	X	M
208		X	X	M
209	X		X	M
210	X		X	M
211		X		C
212	X			C
213	X		X	M
214		X		M
215	X			C
216		X		M
218		X	X	M
219	X		X	M
220		X	X	M
221	X		X	M
222	X		X	M
223	X			M
224	X		X	M
225		X		C
226		X	X	M
227		X		D
228	X			D
229		X	X	M

TABLE 3-VIII. TRAINING AND TEST RESULTS - BY WINDOW (Continued)

<u>Window Number</u>	<u>Training</u>	<u>Test</u>	<u>Degraded?</u>	<u>Result</u>
232		X		M
233	X			M
234	X			C
235		X		C
236		X		M
237	X			C
238	X		X	M
243	X			D
245		X		C
246		X		M
247	X			C
248		X		D
249		X		D
256	X			C
258	X			C
261	X		X	C
263	X			C
265	X		X	M
267	X		X	M

was detected and screened, but misclassified as to target type, a "D" is given. Missed targets are keyed by an "M".

As eluded to earlier, for false alarm testing, non-target windows were lifted from the original digital images the same way as target windows. For the test phase, 213 were processed. After excluding the false alarms caused by extraneous sources, 42 windows or 20% had false alarms. Note that this is lower than the training false alarm rate. For training, windows containing target-like objects were selected, but the test windows were selected to represent areas over the whole field-of-view.

The false alarm rate can be reduced in at least six ways, as follows:

1. Reduce sensitivity threshold
2. Modify classification thresholds
3. Increase prefiltering
4. Tighten detection criteria
5. Use context information
6. Use range information.

As shown in the training results, the rate is lowered considerably by increasing the minimum gradient setting of the preprocessor (reducing the sensitivity). An increase of 1.5 lowered the false alarm rate by 18% in that test. (Par. 3.5). Further insight into the effects of changing the setting can be obtained from the results of a similar test program with FLIR imagery, done for the Army at Frankford Arsenal (Ref. 3). Several hundred samples were run at three different gradient or sensitivity levels. The results are tabulated below.

<u>SENSITIVITY</u>	<u>DETECTION</u>	<u>DETECTED AND SCREENED</u>	<u>DETECTED, SCREENED, CLASSIFIED</u>	<u>FALSE ALARM RATE</u>
Low (4.0)	87%	67%	51%	2%
Medium (3.0)	95%	75%	54%	7%
High (2.0)	98%	85%	58%	19%

The present test results, run at a gradient threshold 2.5, are close to those at the 2.0 setting above. It can be seen that reduction of sensitivity may significantly reduce the false alarm rate while only slightly reducing the final classification rate.

A second option for reducing the rate is in the final processor. If the false alarm closure criterion is increased, the false alarm rate will decrease. As an example, if the minimum acceptable closure is changed from 0.37 to 0.40, the 20% false alarm rate becomes 32 alarms in 213 windows, or 15%. Additional study would be needed to determine how the detection rate is affected by this criterion.

A third method of reducing the false alarms is to increase the amount of prefiltering of the data. Either defocusing-type filtering or more elaborate neighborhood averaging type filtering would reduce those false alarms caused by noise. If the significant target features are not obliterated by the filtering, the recognition rate may be maintained.

False alarms could also be reduced by using only blobs as "cues" to locate candidate objects. Long subsets would not be used as cues. For this set of images, the false alarm count would drop from 42 down to 10 false alarms out of 213 windows, or 5%. Some targets would also be lost, but a modification of the blob detector stage could retrieve a portion of them. Further experiments into this possibility are desirable.

False alarms might also be reduced through the use of texture or context information. Texture statistics are already computed in the preprocessor. Using them to classify the terrain was initially accomplished in the Phase I portion of the Frankford study. Knowledge of the terrain type and use of other background statistics can help prevent false alarms. In the present program, texture statistics were generated. However, training and test efforts to classify the terrain, and incorporation of the data into the decision logic were not within the scope of the program.

Finally, range information can be quite useful in preventing false alarms. Future sensor systems are likely to have available ranging devices for weapon delivery. It was found in the Frankford study that the use of range information aids in rejecting false alarms, as well as increasing target classification accuracy.

Given the variety of possibilities for reducing the false alarm rate, reducing the initial 20% rate to, say, 1% is not unrealistic. Since each window represents 1/80 of the field-of-view area (100 x 100 out of 800 x 1024 pixels), there would be then 0.80 false alarms per frame, or one alarm per 1.25 frames.

To the operator, though, the effective rate would be lower. Except for the occasional effects of noise, new false alarms would ordinarily be generated only as new scenes are covered by the field-of-view. But the scene only slowly changes (over several seconds) when the sensor looks out at targets at long range. Therefore, it should be kept in mind that the false alarm rate per frame will apply to changes of scene in the field-of-view, in the system application, and not to the refresh rate of the sensor. The operator will be faced with one probable false alarm over perhaps 5 to 10 seconds, based on the false alarm rate noted above.

### 3.7 Discussion

At this point, it is timely to note two important points about the limited number of target samples. First, as emphasized by the problem with the truck class, on a per class basis the number of samples may not be sufficient to fully represent their distributions in the feature space. In general, for single modal distributions, a minimum training set should contain at least 10 samples per feature per class to provide a suitable estimate of the distribution. (Although in practice, that is often difficult to achieve.) In this case, there are 10 different features employed in the classification boundaries. So while approximately 100 truck samples would be desirable, only 19 were available for training. This naturally creates difficulties in estimating suitable boundaries for adequate performance on completely new samples (e.g., the test set).

The second point concerning the number of samples is the confidence level. The statistical nature of the test creates some uncertainty about the performance estimates. The confidence interval expresses how much confidence is justified in the sample set. For example, the test score for Detected, Screened, and Classified samples was 48%. If we assume that the outcome of this score was binomially distributed\* - a yes or no scoring, then the 95% confidence interval for 50 samples is 33% to 63%. Additional samples would narrow this wide range.

For a more "averaged" look at the performance estimates, the training and test results are combined in Table 3-IX. The degraded images have been

---

\*A questionable assumption in view of the complexity of the features and classification algorithm. So this is likely to be an optimistic estimate.

### 3.7 Discussion

At this point, it is timely to note two important points about the limited number of target samples. First, as emphasized by the problem with the truck class, on a per class basis the number of samples may not be sufficient to fully represent their distributions in the feature space. In general, for single modal distributions, a minimum training set should contain at least 10 samples per feature per class to provide a suitable estimate of the distribution. (Although in practice, that is often difficult to achieve.) In this case, there are 10 different features employed in the classification boundaries. So while approximately 100 truck samples would be desirable, only 19 were available for training. This naturally creates difficulties in estimating suitable boundaries for adequate performance on completely new samples (e.g., the test set).

The second point concerning the number of samples is the confidence level. The statistical nature of the test creates some uncertainty about the performance estimates. The confidence interval expresses how much confidence is justified in the sample set. For example, the test score for Detected, Screened, and Classified samples was 48%. If we assume that the outcome of this score was binomially distributed\* - a yes or no scoring, then the 95% confidence interval for 50 samples is 33% to 63%. Additional samples would narrow this wide range.

For a more "averaged" look at the performance estimates, the training and test results are combined in Table 3-IX. The degraded images have been

---

\*A questionable assumption in view of the complexity of the features and classification algorithm. So this is likely to be an optimistic estimate.

excluded since they are probably unrealistic in terms of pure FLIR video. Detection, and Classification are both good. The "product" score Detected, Screened, and Classified is also acceptable.

To put the scores into perspective, consider the performance of a human interpreter. The Detected and Screened score, as previously noted, is equivalent to what a human observer would call "detection", and the classification of Detected and Screened score is equivalent to an observers "recognition" rate. Mr. John Dehne of NVL indicates that for this type of imagery, an observer's detection rate is approximately 90% and the recognition rate is about 50%, under ideal conditions.

Additionally, an adhoc experiment was performed in-house on this particular set of imagery. The experiment was carried out with a volunteer subject\* who had not studied or viewed separately the training and test sets. Three different sets of 50 x 50 windows (gray level playbacks) were viewed and classified by the subject. At the end of each of the three sets, a score was made, allowing some feedback to the subject. The totals of the three tests are shown in Table 3-X. Even here, the truck class was inferior. The average score was only 60%, indicating difficulty with this imagery base. This score is the "equivalent" of the Classification of Detected and Screened Samples because the subject knew that every sample viewed did contain a target. The false alarm or detection rate was not investigated. It is not intended to imply that the machine classifier is better than the human - the conditions were not identical and the number of samples too limited. However, it is gratifying to note that they are not highly different.

---

\*Actually, one of the program members, who had had only an initial acquaintance with the imagery.

TABLI: 3 - X

INTERPRETER PERFORMANCE IN RECOGNITION

	<u>TANK</u>	<u>TRUCK</u>	<u>APC</u>	<u>TOTAL</u>
NUMBER OF SAMPLES	43	26	30	99
CORRECTLY CLASSIFIED	31	11	17	59
% CORRECT CLASSIFICATION	72	42	57	60

#### 4.0 CONCLUSIONS AND RECOMMENDATIONS

The objective of this study was to estimate the performance of the digital image processing techniques that have been developed on imagery from an 875-line TV compatible FLIR sensor. Conclusions derived from this study are discussed below. In addition, recommendations are made with regard to the emphasis of future efforts.

From the simulation test described in Section 3.0, the following conclusions are drawn:

1. Initial acquisition of target material is in the 90% range.  
However, rejection of some targets is necessary to limit the rate of false alarms. The best compromise depends upon mission requirements.
2. Classification performance was generally in the 60-80% range.  
Specific performance depended upon the size of the training set and the quality of the images.
3. Extraneous sources of degradation of the imagery made testing and evaluation more difficult. In practice, it is assumed that most of these sources would not be present in the FLIR video.
4. In view of the large number of features needed to separate the target classes, the size of the data base was very limited. This made extrapolation from the training samples to the test samples a precarious trial for the classifier.

#### 4.0 CONCLUSIONS AND RECOMMENDATIONS

The objective of this study was to estimate the performance of the digital image processing techniques that have been developed on imagery from an 875-line TV compatible FLIR sensor. Conclusions derived from this study are discussed below. In addition, recommendations are made with regard to the emphasis of future efforts.

From the simulation test described in Section 3.0, the following conclusions are drawn:

1. Initial acquisition of target material is in the 90% range.  
However, rejection of some targets is necessary to limit the rate of false alarms. The best compromise depends upon mission requirements.
2. Classification performance was generally in the 60-80% range.  
Specific performance depended upon the size of the training set and the quality of the images.
3. Extraneous sources of degradation of the imagery made testing and evaluation more difficult. In practice, it is assumed that most of these sources would not be present in the FLIR video.
4. In view of the large number of features needed to separate the target classes, the size of the data base was very limited. This made extrapolation from the training samples to the test samples a precarious trial for the classifier.

## 5.0 REFERENCES

1. "Automatic Target Cueing Study", Contract N00019-71-C-0253 to the Naval Air Systems Command.
2. "Advanced Analysis of SLR Changes", Contract F-30602-72-C-0126 to Rome Air Development Center.
3. "Automatic Target Cueing Study for Helicopter Fire Control", Contract DAAA25-73-C-0719 to Frankford Arsenal.
4. "Recognition System Design By Statistical Analysis", L. Kanal and N. Randall, from Proceedings of the National Conference of the ACM, 1964.